

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: October 5, 2005, 07:48:48 : Search time 183.331 Seconds  
(without alignments)  
2018.911 Million cell updates/sec

Title: US-09-917-376-1

Perfect score: 5135

Sequence: 1 MRSSENIRLTMRSRLVSL.....RASFGSVNPRATPYLQX 957

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq.16Dec04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5134	100.0	957	6	ABP73015
2	5134	100.0	957	7	ADD22921
3	5134	100.0	957	8	ADO52313
4	4036	78.6	740	6	ABP73016
5	4036	78.6	740	7	ADD22926
6	4036	78.6	740	7	ADD22923
7	3920.5	76.3	733	8	ADO52315
8	2174	42.3	940	6	ABB99489
9	1680	32.7	726	7	ADD22927
10	1625.5	31.7	838	7	ADD42055
11	1588	30.9	818	8	ADH51581
12	1125	21.9	812	8	ADH15118
13	1117	21.8	789	8	ADH19120
14	1114	21.7	826	8	ADH19124
15	993	19.3	776	8	ADR90296
16	979	19.1	756	8	ADR90298
17	979	19.1	757	8	ADR90302
18	848	16.5	555	8	ADD24919
19	823	16.0	1228	6	ABP73029
20	823	16.0	1228	6	ADH36363
21	702	13.7	762	8	ABP73022
22	702	13.7	762	7	ADJ38291
23	468	9.1	88	6	ABP73018
24	468	9.1	88	8	ADO52317
25	468	9.1	89	6	ABP73017

26	468	9.1	89	8	ADO52316	Ado52316 A. cellu
27	462	9.0	88	7	ADD22925	Add22925 Acidother
28	462	9.0	89	7	ADD22924	Add22924 Acidother
29	462	9.0	154	6	ABP73020	Abp73020 Amino aci
30	462	9.0	154	6	ADJ38294	Adj38294 A. cellu
31	459	8.9	150	6	ABP73025	Abp73025 Amino aci
32	459	8.9	150	7	ADH36640	Adh36640 A. cellu
33	431	8.4	1121	8	ADP75895	Adp75895 Acidother
34	429	8.4	153	6	ADP71658	Adp71658 A. cellu
35	429	8.4	153	7	ADP75898	Adp75898 Carbohyd
36	429	8.4	1121	6	ABP71656	Abp71656 A. cellu
37	374	6.5	1751	5	AAE13493	Aay13493 Truncated
38	334	6.5	1751	5	AAE16324	Aae16324 Active ce
39	323	6.3	782	2	AAE15625	Aae15625 Glutamic
40	301.5	5.9	1749	4	AAE72314	Aae72314 Glutamic
41	288	5.6	96	5	ABB78539	Abb78539 Gum arabi
42	276	5.4	1426	2	AAE13492	Aay13492 Truncated
43	276	5.4	1426	5	AAE16323	Aae16323 Active ce
44	274	5.3	406	4	ABG27250	Abg27250 Novel hum
45	262	5.1	2468	6	ABU38411	Abu38411 Protein e

#### ALIGNMENTS

##### RESULT 1

ID ABP73015 standard; protein; 957 AA.

AC ABP73015;

DT 03-JUN-2003 (first entry)

DE Amino acid sequence of the avicelase AvIII.

KW Avicelase; AvIII; glycoside hydrolase; enzyme; cellulase; biofuel;

KW detergent; pulp processing; paper processing; feed processing; textile;

KW cellulose.

OS Acidothermus cellulosoliticus.

XX WO2003012090-A2.

XX 13-FEB-2003.

XX 28-JUL-2001; 2001WO-US023818.

XX 28-JUL-2001; 2001WO-US023818.

XX (MIDE ) MIDWEST RES INST.

XX Ding S, Adney WS, Vinzant TB, Himmel ME;

XX WPI; 2003-248177/24.

XX N-PSDB; ABE77632.

XX New thermostable AvIII peptide from Acidothermus cellulosoliticus, useful

XX for degradation of cellulose or in generating anti-AvIII antibodies for

XX purifying recombinant AvIII polypeptides from genetically engineered

XX host cells.

XX Claim 2; Page 20; 44pp; English.

XX The present sequence represents a thermostable avicelase polypeptide,

XX designated AvIII. AvIII is a member of the glycoside hydrolase family

XX of enzymes, and is a cellulase. AvIII is useful in the conversion of

XX biomass to biofuels and biofuel additives. It may be useful in the

XX production of detergents, pulp and paper processing, food and feed

XX processing and in textile processes. The thermostable AvIII peptide is



CC useful in the degradation of cellulose, and in generating specific anti-AviIII antibodies that are useful in purifying recombinant AviIII polypeptides from genetically engineered host cells, in detecting AviIII polypeptide expression, as well as a reagent tool for characterizing the molecular actions of the polypeptide. The AviIII polynucleotide is useful as a source of probes or primers in various diagnostic assays

XX Sequence 957 AA;

Query Match 100.0%; Score 5134; DB 6; Length 957;  
Best Local Similarity 100.0%; Pred. No. 4.3e-263;  
Matches 956; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDRSENIRLITMRSRRRLVSLAATASFAVAALGVLPLAITASPAHAATTOPTYSNVAIG 60  
DB 1 MDRSENIRLITMRSRRRLVSLAATASFAVAALGVLPLAITASPAHAATTOPTYSNVAIG 60  
QY 61 GGGFVDGIVFNEGAPGILLYRTDIGMYRMDANGRWIPLLDWVGNNMGVNGVSIAD 120  
DB 61 GGGFVDGIVFNEGAPGILLYRTDIGMYRMDANGRWIPLLDWVGNNMGVNGVSIAD 120  
QY 121 PINTNKVMAVGMTNNSWDPNDGAILRSSDQATWQITPLPFLKGGNMPGKMGRLAVD 180  
DB 121 PINTNKVMAVGMTNNSWDPNDGAILRSSDQATWQITPLPFLKGGNMPGKMGRLAVD 180  
QY 181 PNNNDILYFGAPGKGLMRTSGATWQMTNPDVCTYIANPTDTGYQSDIQGVVWA 240  
DB 181 PNNNDILYFGAPGKGLMRTSGATWQMTNPDVCTYIANPTDTGYQSDIQGVVWA 240  
QY 181 PNNNDILYFGAPGKGLMRTSGATWQMTNPDVCTYIANPTDTGYQSDIQGVVWA 240  
DB 181 PNNNDILYFGAPGKGLMRTSGATWQMTNPDVCTYIANPTDTGYQSDIQGVVWA 240  
QY 241 FPKSSSLGQASRTIFVGVADPNNPFWMSRDGATQAVGAPLPGFIPHKGVDPVNVHL 300  
DB 241 FPKSSSLGQASRTIFVGVADPNNPFWMSRDGATQAVGAPLPGFIPHKGVDPVNVHL 300  
QY 301 YIATSNVTGPRYDSSGDVWKFVTSQWTRISVPSTDTANDYFGVSGLTIDRQHPNTIM 360  
DB 301 YIATSNVTGPRYDSSGDVWKFVTSQWTRISVPSTDTANDYFGVSGLTIDRQHPNTIM 360  
QY 361 VATQISWMPDITIFRSTDGATWTRIMDWTSPNRSRLRYLDISAEPMLEFGVQPNPVP 420  
DB 361 VATQISWMPDITIFRSTDGATWTRIMDWTSPNRSRLRYLDISAEPMLEFGVQPNPVP 420  
QY 421 SPRLGMDAMALIDPNSDRMLYGTATLYATNDLTKWDSGGQIHIAPMKGLLEETVND 480  
DB 421 SPRLGMDAMALIDPNSDRMLYGTATLYATNDLTKWDSGGQIHIAPMKGLLEETVND 480  
QY 481 LISPPGAPLISALGDLGCTHADYTAVPSTIFTSPVFTTGSTVYALNPSIIVRAGSF 540  
DB 481 LISPPGAPLISALGDLGCTHADYTAVPSTIFTSPVFTTGSTVYALNPSIIVRAGSF 540  
QY 541 DPSSQPNDRHVAFTSDGKNWFOGSEPGVTTGCTVAASADGSRFVWAPDPCQPVVAV 600  
DB 541 DPSSQPNDRHVAFTSDGKNWFOGSEPGVTTGCTVAASADGSRFVWAPDPCQPVVAV 600  
QY 601 GPGNSMAAOGVANNQISDRNPKTFYALSNCTFPRSDGVTTPQVAAAGLPSSGAVG 660  
DB 601 GPGNSMAAOGVANNQISDRNPKTFYALSNCTFPRSDGVTTPQVAAAGLPSSGAVG 660  
QY 661 VNFHAVPKGEGDMLAASGLYHSTNGSSMSAITGVSSAVNVFPGSAPGSSPYPAFVV 720  
DB 661 VNFHAVPKGEGDMLAASGLYHSTNGSSMSAITGVSSAVNVFPGSAPGSSPYPAFVV 720  
QY 721 GTTGGVTGAYRSDDCGTTWVLINDQHOYGNMGOAITGDHANLRVYIGTNGRCIYVDI 780  
DB 721 GTTGGVTGAYRSDDCGTTWVLINDQHOYGNMGOAITGDHANLRVYIGTNGRCIYVDI 780  
QY 781 GGAPSSSPSPSVSPASPSLSBPSPSSSPSSPSSSPSSSPSSPSPSPSPSPSPSPSPSP 840  
DB 781 GGAPSSSPSPSVSPASPSLSBPSPSSSPSSPSSSPSSSPSSPSPSPSPSPSPSPSPSP 840  
QY 841 SASPSSSPSPSSPSSP 900  
DB 841 SASPSSSPSPSSPSSP 900

QY 901 SVDLSVTYVRYWFTRDGSSSTLYVNCDMAIGGNIRASFGSVNPAITPATYLIQ 956  
DB 901 SVDLSVTYVRYWFTRDGSSSTLYVNCDMAIGGNIRASFGSVNPAITPATYLIQ 956

# RESULT 2

ADD22921  
ID ADD22921 standard; protein; 957 AA.

XX AC ADD22921;

DT 15-JAN-2004 (first entry)

XX

DE Acidohermus cellulolyticus avicelase AviIII.

KM enzyme; AviIII; cellulose reduction; agricultural biomass;

XX municipal solid waste; glycoside hydrolase; avicelase.

OS Acidohermus cellulolyticus.

XX

FT Key Location/Qualifiers

FT Misc-difference 957 /label= Unknown

FT /note= "Encoded by N"

XX

PN US2003108988-A1.

XX

PD 12-JUN-2003.

XX

PF 18-OCT-2002; 2002US-0015400.

XX

PR 28-JUL-2001; 2001US-00917376.

XX

PA (DING/) DING S.

XX

PA (ADNE/) ADNEY W S.

XX

PA (VINZ/) VINZANT T B.

XX

PA (HIMM/) HIMMEL M E.

XX

P1 Ding S, Adney WS, Vinzant TB, Himmel ME;

XX

DR WPI; 2003-810853/76.

XX

DR N-PSDB; ADD22922.

XX

PT New isolated thermal tolerant avicelase polynucleotide useful for detection of a polynucleotide encoding AviIII and for reducing cellulose in a starting material, e.g. municipal solid waste.

XX

PS Claim 16; SEQ ID NO 1; 29pp; English.

XX

XX The invention relates to an isolated polynucleotide molecule encoding a

CC

CC thermostable AviIII polypeptide. The polynucleotide is useful for

CC

CC detection of a polynucleotide encoding AviIII. The polynucleotide is

CC

CC useful for reducing cellulose in a starting material which involves

CC

CC administering to the starting material, e.g. agricultural biomass or

CC

CC municipal solid waste, a polypeptide molecule of the polynucleotide. The

CC

CC method further comprises administering a second polypeptide molecule

CC

CC chosen from the glycoside hydrolase family of proteins. The present

CC

CC sequence represents the amino acid sequence of Acidohermus

XX

XX cellulolyticus avicelase AviIII.

XX Sequence 957 AA;

Query Match 100.0%; Score 5134; DB 7; Length 957;

Best Local Similarity 100.0%; Pred. No. 4.3e-263;

Matches 956; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDRSENIRLITMRSRRRLVSLAATASFAVAALGVLPLAITASPAHAATTOPTYSNVAIG 60  
DB 1 MDRSENIRLITMRSRRRLVSLAATASFAVAALGVLPLAITASPAHAATTOPTYSNVAIG 60  
QY 61 GGGFVDGIVFNEGAPGILLYRTDIGMYRMDANGRWIPLLDWVGNNMGVNGVSIAD 120  
DB 61 GGGFVDGIVFNEGAPGILLYRTDIGMYRMDANGRWIPLLDWVGNNMGVNGVSIAD 120



```

QY 121 PINTNKVAAVGMVYTNMSPNDGAILRSSDQATWQITPLPFLKGGNMPGRGMBRLAVD 180
DB 121 PINTNKVAAVGMVYTNMSPNDGAILRSSDQATWQITPLPFLKGGNMPGRGMBRLAVD 180
QY 181 PANNILYFGAPSGKGLMRSTDGATWQMTNFPDVGTIYANPTDTTGYOSDIQGVVWA 240
DB 181 PANNILYFGAPSGKGLMRSTDGATWQMTNFPDVGTIYANPTDTTGYOSDIQGVVWA 240
QY 241 FDKSSSSILGOASKTITFVGVADPNNPVFWSRDGATWQAVGAPGTGFIPIHKGVDPVNVHL 300
DB 241 FDKSSSSILGOASKTITFVGVADPNNPVFWSRDGATWQAVGAPGTGFIPIHKGVDPVNVHL 300
QY 301 YIATSNITGPGYDSSGDVWKFVSSTGVTWRIISPVPSTDTANDYFGVSGLTIDRQHPNTIM 360
DB 301 YIATSNITGPGYDSSGDVWKFVSSTGVTWRIISPVPSTDTANDYFGVSGLTIDRQHPNTIM 360
QY 361 VATQISWMPDITIIFRSTDGATWTRIMDMTSPNRSIRVYLDISAEFWLTFGVQPNPVP 420
DB 361 VATQISWMPDITIIFRSTDGATWTRIMDMTSPNRSIRVYLDISAEFWLTFGVQPNPVP 420
QY 421 SPKLGMDAMALDPFNSDRLVGTGATLYATNDLYKMSGGQIHIAPMYKGLBETAVND 480
DB 421 SPKLGMDAMALDPFNSDRLVGTGATLYATNDLYKMSGGQIHIAPMYKGLBETAVND 480
QY 481 LISPPSGALISALGDLGGFTHADVTAVPSTIFTSPTTGTVDVAELNPSIIVRAGSF 540
DB 481 LISPPSGALISALGDLGGFTHADVTAVPSTIFTSPTTGTVDVAELNPSIIVRAGSF 540
QY 541 DSSSQNDNRHVAESTDGGKMFQSGSEPGVTTGTVASAADGRFVWAPDPCQPVYAV 600
DB 541 DSSSQNDNRHVAESTDGGKMFQSGSEPGVTTGTVASAADGRFVWAPDPCQPVYAV 600
QY 601 GFNSWVAASQGVPAQAQIRSDRVNPKTFYALSNGTFRSTDGVTGQVPAAGLPSSGAVG 660
DB 601 GFNSWVAASQGVPAQAQIRSDRVNPKTFYALSNGTFRSTDGVTGQVPAAGLPSSGAVG 660
QY 661 VMFHAVPGKEGDIWLAASGLYHSTNGSSWSAITGVSAVNVGFGKSAAGSSYPVAVV 720
DB 661 VMFHAVPGKEGDIWLAASGLYHSTNGSSWSAITGVSAVNVGFGKSAAGSSYPVAVV 720
QY 721 GTTIGTVAGYRSDCCGTTWLIINDDOQYGNMGQAITGDHANLRVYIGTNGRGIYVDI 780
DB 721 GTTIGTVAGYRSDCCGTTWLIINDDOQYGNMGQAITGDHANLRVYIGTNGRGIYVDI 780
QY 781 GGAPSGSPSPSVASPSLSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 840
DB 781 GGAPSGSPSPSVASPSLSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 840
QY 841 SASPSPSSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 900
DB 841 SASPSPSSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 900
QY 901 SSVYDLSVTVYRMYFTIRBGSSSTLYVNCDMAIIGCGNIRASFGSVNPAITPADIYLO 956
DB 901 SSVYDLSVTVYRMYFTIRBGSSSTLYVNCDMAIIGCGNIRASFGSVNPAITPADIYLO 956

```

```

XX Key Location/Qualifiers
FH Peptide 1..34
FT /label= signal_peptide
FT Protein 35..957
FT Domain 35..187
FT Domain /note= "Carbohydrate binding domain type III"
FT Domain 231..870
FT /note = GH74 catalytic domain
FT Misc-difference 957
FT /label= Unknown
XX US2004038334-A1.
XX 26-FEB-2004.
XX 28-JUL-2001; 2001US-00917376.
XX 28-JUL-2001; 2001US-00917376.
XX PR 28-JUL-2001; 2001US-00917376.
XX PA (DING/) DING S.
XX PA (ADNEY/) ADNEY W S.
XX PA (VINZ/) VINZANT T B.
XX PA (HIMM/) HIMMEL M E.
XX PI Ding S, Adney WS, Vinzant TB, Himmel ME;
XX DR WPI; 2004-203224/19.
XX DR N-PSDB; ADO52314.
XX PT Novel thermostable Avitrii polypeptide of glycoside hydrolase family and
PT isolated from Acidothermus cellulolyticus, useful for degrading treated
PT biomass into simpler forms of carbohydrate.
XX PS Claim 12; SEQ ID NO 1; 19pp; English.
XX CC The invention relates to a thermostable cellulase enzyme, avicelase III
CC (Avitrii) and its nucleic acid sequence. Avitrii is useful for reducing
CC cellulose in a starting material. A thermostable Avitrii peptide is useful
CC for degrading treated biomass into simpler forms of carbohydrate, which
CC is used in the formation of ethanol or other industrial chemicals. It is
CC also useful for treating fabrics to remove cellulose-containing stains.
CC The present sequence is Acidothermus cellulolyticus Avitrii protein.
XX SQ Sequence 957 AA:
XX Query Match 100.0%; Score 5134; DB 8; Length 957;
XX Best Local Similarity 100.0%; Pred. No. 4.3e-263;
XX Matches 956; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MDRESEIRLTVRSRLVSLAATASFAVAALGVPIAITASPAHAATTQPTYSNVAIG 60
DB 1 MDRESEIRLTVRSRLVSLAATASFAVAALGVPIAITASPAHAATTQPTYSNVAIG 60
QY 61 GGGFYDGIYVNEGAPGILTYRDTICGMYRMDAANGRWLPLDWGMNMGNGVVSIAAD 120
DB 61 GGGFYDGIYVNEGAPGILTYRDTICGMYRMDAANGRWLPLDWGMNMGNGVVSIAAD 120
QY 121 PINTNKVAAVGMVYTNMSPNDGAILRSSDQATWQITPLPFLKGGNMPGRGMBRLAVD 180
DB 121 PINTNKVAAVGMVYTNMSPNDGAILRSSDQATWQITPLPFLKGGNMPGRGMBRLAVD 180
QY 181 PANNILYFGAPSGKGLMRSTDGATWQMTNFPDVGTIYANPTDTTGYOSDIQGVVWA 240
DB 181 PANNILYFGAPSGKGLMRSTDGATWQMTNFPDVGTIYANPTDTTGYOSDIQGVVWA 240
QY 241 FDKSSSSILGOASKTITFVGVADPNNPVFWSRDGATWQAVGAPGTGFIPIHKGVDPVNVHL 300
DB 241 FDKSSSSILGOASKTITFVGVADPNNPVFWSRDGATWQAVGAPGTGFIPIHKGVDPVNVHL 300
QY 301 YIATSNITGPGYDSSGDVWKFVSSTGVTWRIISPVPSTDTANDYFGVSGLTIDRQHPNTIM 360
DB 301 YIATSNITGPGYDSSGDVWKFVSSTGVTWRIISPVPSTDTANDYFGVSGLTIDRQHPNTIM 360

```



Db	301	VIATSNITGGPXYDGGSSGDWVKFVSITGSTWTRISRPSTVDIANDVFGVSGGLTIDQGHNTIM	36
QY	361	VATQISMPDPTIIERSTDCGATWTRIMWTSTYPNRSLRYLIDISABPMLTFEQPNPVP	420
Db	361	VATQISMPDPTIIFRSTDCGATWTRIMWTSTYPNRSLRYLIDISABPMLTFEQPNPVP	420
QY	421	SPKIGMDMEAMIDRPNSDRMLYTGATLYATNDLTKMDSGGQIHIAPMVKGLEFNAVND	480
Db	421	SPKIGMDMEAMIDRPNSDRMLYTGATLYATNDLTKMDSGGQIHIAPMVKGLEFNAVND	480
QY	481	LISPSGAPLISALGDLGFTADVAVAVSTIPTSSEVFTTGTVDVAELNPSIIVRAGSF	540
Db	481	LISPSGAPLISALGDLGFTADVAVAVSTIPTSSEVFTTGTVDVAELNPSIIVRAGSF	540
QY	541	DPSGQPNDRHVAFSTDGGKNWFGQSEBPGVTTGTVAASADGSRFYWAPGDPQPVYVAV	600
Db	541	DPSGQPNDRHVAFSTDGGKNWFGQSEBPGVTTGTVAASADGSRFYWAPGDPQPVYVAV	600
QY	601	GFQNSMAASQVYPANAQIISDRVNPRTFYALSNGTYYRSTIDGVTTPQVAAGLPSSGAVG	660
Db	601	GFQNSMAASQVYPANAQIISDRVNPRTFYALSNGTYYRSTIDGVTTPQVAAGLPSSGAVG	660
QY	661	VMFHAVPGKEGDLMLTAASSGLYHSTTNGGSSMSAITVSSASVANYGFGKSAPGSSYPAVFYV	720
Db	661	VMFHAVPGKEGDLMLTAASSGLYHSTTNGGSSMSAITVSSASVANYGFGKSAPGSSYPAVFYV	720
QY	721	GTIGGVTGAYRSDDCGTTWVLINDDQHOYGNMGOALITGDHANLRVYIGTNGRGIVYGD1	780
Db	721	GTIGGVTGAYRSDDCGTTWVLINDDQHOYGNMGOALITGDHANLRVYIGTNGRGIVYGD1	780
QY	781	GCAPSGSGSPSVSPASPELSPSPSSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSP	840
Db	781	GCAPSGSGSPSVSPASPELSPSPSSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSP	840
QY	841	SASPSPPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSP	900
Db	841	SASPSPPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSP	900
QY	901	SSVDLSTVYVRWFTFRDGGSSTLVYNCDMAALICGNIIRASFGSVNPATPTADTYLQ	956
Db	901	SSVDLSTVYVRWFTFRDGGSSTLVYNCDMAALICGNIIRASFGSVNPATPTADTYLQ	956
RESULT 4			
ABP73016			
ID	ABP73016	standard; peptide; 740 AA.	
XX	ABP73016;		
AC			
XX			
XX			
DT	03-JUN-2003 (first entry)		
XX			
DE	Amino acid sequence of the avicelase AvIII catalytic domain.		
KM	Avicelase; AvIII; glycoside hydrolase; enzyme; cellulase; bifuel;		
KW	detergent; pulp processing; paper processing; feed processing; textile;		
KM	cellulose.		
XX			
OS	Acidothermus cellulolyticus.		
XX			
FN	WO2003012090-A2.		
PD	13-FEB-2003.		
XX			
XX			
PF	28-JUL-2001; 2001WO-US023818.		
XX			
FR	28-JUL-2001; 2001WO-US023818.		
XX			
PA	(MIDE ) MIDWEST RES INST.		
PI	Ding S, Adney WS, Vinzant TB, Himmel ME;		
XX			
DR	WPI; 2003-248177/24.		

PT	New thermostable AviIII peptide from <i>Acidothermus cellulolyticus</i> , useful
PT	for degradation of cellulose or in generating anti-AviIII antibodies for
PT	purifying recombinant AviIII polypeptides from genetically engineered
PT	host cells.
XX	
PS	Claim 6; Page 8; 44pp; English.
CC	
CC	The present sequence is derived from a thermostable avicelase, designated
CC	AviIII. AviIII is a member of the glycoside hydrolase family of enzymes,
CC	and is a cellulase. AviIII is useful in the conversion of biomass to
CC	biofuels and biocatal additives. It may be useful in the production of
CC	detergents, pulp and paper processing, food and feed processing and in
CC	textile processes. The thermostable AviIII peptide is useful in the
CC	degradation of cellulose, and in generating specific anti-AviIII
CC	antibodies that are useful in purifying recombinant AviIII polypeptides
CC	from genetically engineered host cells, in detecting AviIII polypeptide
CC	expression, as well as a reagent tool for characterizing the molecular
CC	actions of the polypeptide. The AviIII polynucleotide is useful as a
CC	source of probes or primers in various diagnostic assays
XX	
SQ	Sequence 740 AA;
Query Match	78.6%; Score 4036; DB 6; Length 740;
Best Local Similarity	100.0%; Pred. No. 3,66-205;
Matches 740; Conservative	0; Mismatches 0; Indels 0; Gaps 0
DB	1 ATGQYTTSMNAVIGGGGFGVDIGVFNENAGPILVYRTDIGMYRMDANGRWIPLLDWGM 60
QY	47 ATGQYTTSMNAVIGGGGFGVDIGVFNENAGPILVYRTDIGMYRMDANGRWIPLLDWGM 106
QY	107 NNMGNGVNSTAADPINTNKWMAAVGMTTNSMDPNDGAILRSSDGAATWQITPLPEPKLGG 166
DB	61 NNMGNGVNSTAADPINTNKWMAAVGMTTNSMDPNDGAILRSSDGAATWQITPLPEPKLGG 120
QY	167 NMPGRGMGERLAVDNNNNILYFGAPSGKGLMRSTDSGATSGMTNPFVGYIYANPDT 226
DB	121 NMPGRGMGERLAVDNNNNILYFGAPSGKGLMRSTDSGATSGMTNPFVGYIYANPDT 180
QY	227 TGYQSDIGVWVAFAFDKSSSSISGASAKTIPVGVADPNPVPFMSRDGATWQAVGAPTF 286
DB	181 TGYQSDIGVWVAFAFDKSSSSISGASAKTIPVGVADPNPVPFMSRDGATWQAVGAPTF 240
QY	287 IPHKGVDPVNHVLIATISNTGGPYDSSGDVWKRSTVSGTTRISLPPSTDTANDYFEG 346
DB	241 IPHKGVDPVNHVLIATISNTGGPYDSSGDVWKRSTVSGTTRISLPPSTDTANDYFEG 300
QY	347 SGLTIDRQHPNTIMATQISWMPDITIRSTDDGATWTRIMWTSYPNRSRLRYVDISAE 406
DB	301 SGLTIDRQHPNTIMATQISWMPDITIRSTDDGATWTRIMWTSYPNRSRLRYVDISAE 360
QY	407 PWLTFGVDPNPVPPPKLGMNDEAALIDPFNSDRMLYTGATLYATNDLTTKMDSGGQIHI 466
DB	361 PWLTFGVDPNPVPPPKLGMNDEAALIDPFNSDRMLYTGATLYATNDLTTKMDSGGQIHI 420
QY	467 APMVKGLBETAVNDLISPPSGAPLISALGDLGFTHADYAVPSTIFTSVPFTGTSDY 526
DB	421 APMVKGLBETAVNDLISPPSGAPLISALGDLGFTHADYAVPSTIFTSVPFTGTSDY 480
QY	527 AEINSSIIYRAGSPSPSQPNDRHVAFTSDGKKMFGQSEPGQVTTGGVVAASADGSRV 586
DB	481 AEINSSIIYRAGSPSPSQPNDRHVAFTSDGKKMFGQSEPGQVTTGGVVAASADGSRV 540
QY	587 WAPGPGQGVVYAVFGNSMAASQGVPAQAQIRSRVNPVKTETALNSNGTFYASTDGVTF 646
DB	541 WAPGPGQGVVYAVFGNSMAASQGVPAQAQIRSRVNPVKTETALNSNGTFYASTDGVTF 600
QY	647 QPVAAGLPSSGAVGMFHAVPGKEDLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 706
DB	601 QPVAAGLPSSGAVGMFHAVPGKEDLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 660
QY	707 KSAPGSSIPAVNVGTIGGVTGAYRSDCGTTWVLINDQHQYGMWGAITGDHNLRLRV 766
DB	661 KSAPGSSIPAVNVGTIGGVTGAYRSDCGTTWVLINDQHQYGMWGAITGDHNLRLRV 720



QY 767 YIGTNGRGIVGDIIGAPSG 786  
 DB 721 YIGTNGRGIVGDIIGAPSG 740

RESULT 5  
 ADD22926  
 ID ADD22926 standard; protein; 740 AA.  
 AC ADD22926;  
 XX 15-JAN-2004 (first entry)  
 DT  
 XX Acidothermus cellulolyticus avicelase AvIII catalytic domain.  
 DE  
 XX enzyme; AvIII; cellulose reduction; agricultural biomass;  
 KM municipal solid waste; glycoside hydrolase; avicelase.  
 XX  
 OS Acidothermus cellulolyticus.  
 XX  
 XX US2003108988-A1.  
 XX  
 PD 12-JUN-2003.  
 XX  
 PF 18-OCT-2002; 2002US-00155400.  
 XX  
 PR 28-JUL-2001; 2001US-00917376.  
 XX  
 XX (DING/) DING S.  
 PA (ADNEY/) ADNEY W S.  
 PA (VINZ/) VINZANT T B.  
 PA (HIMM/) HIMMEL M E.  
 XX  
 PI Ding S, Adney WS, Vinzant TB, Himmel ME;  
 DR WPI; 2003-810853/76.  
 XX  
 PT New isolated thermal tolerant avicelase polynucleotide useful for  
 PT detection of a polynucleotide encoding AvIII and for reducing cellulose  
 PT in a starting material, e.g. municipal solid waste.  
 XX  
 PS Example 2; SEQ ID NO 6; 29pp; English.  
 XX  
 CC The invention relates to an isolated polynucleotide molecule encoding a  
 CC thermostable AvIII polypeptide. The polynucleotide is useful for  
 CC detection of a polynucleotide encoding AvIII. The polynucleotide is  
 CC useful for reducing cellulose in a starting material which involves  
 CC administering to the starting material, e.g. agricultural biomass or  
 CC municipal solid waste, a polypeptide molecule of the polynucleotide. The  
 CC method further comprises administering a second polypeptide molecule  
 CC chosen from the glycoside hydrolase family of proteins. The present  
 CC sequence represents the amino acid sequence of Acidothermus  
 CC cellulolyticus avicelase AvIII catalytic domain.  
 CC  
 SQ Sequence 740 AA;

Query Match 78.6%; Score 4036; DB 7; Length 740;  
 Best Local Similarity 100.0%; Pred. No. 3.6e-205; Indels 0; Gaps 0;  
 Matches 740; Conservative 0; Mismatches 0;

QY 47 ATTQPTWSNVAIGGGFVDGIVNEGAPGILYVRTDIGMYRMDAANGRIPLLDWGW 106  
 DB 1 ATTQPTWSNVAIGGGFVDGIVNEGAPGILYVRTDIGMYRMDAANGRIPLLDWGW 60

QY 107 NNGYNGVSVIAADPINTNKVMAAVGYTNSWDPNDGAILRSSDQATWOTPLPFGKGG 166  
 DB 61 NNGYNGVSVIAADPINTNKVMAAVGYTNSWDPNDGAILRSSDQATWOTPLPFGKGG 120

QY 167 NMGRGNGERLAVDPNNNDILYFGAPSGKLMRSTDSGATWSOMTNPFDVGYIANPTDT 226  
 DB 121 NMGRGNGERLAVDPNNNDILYFGAPSGKLMRSTDSGATWSOMTNPFDVGYIANPTDT 180

QY 227 TGYQSDIQGVVWMAFDKSSSSSLGQASKTIFVGVADPNNPVFWSRDQGTWQAVPGAPTYGE 286  
 DB 181 TGYQSDIQGVVWMAFDKSSSSSLGQASKTIFVGVADPNNPVFWSRDQGTWQAVPGAPTYGE 240

QY 287 IPHKGVDPPVNHVLYIATSNVTGSPYDSSGGVWKEFESTTSGTWTRISLPPSTDTANDYGY 346  
 DB 241 IPHKGVDPPVNHVLYIATSNVTGSPYDSSGGVWKEFESTTSGTWTRISLPPSTDTANDYGY 300

QY 347 SGLTIDROHPNTIWAQISWMPDTIIFRSTIDGATWTRIMWMTSYPRRSIRYVLIDISAE 406  
 DB 301 SGLTIDROHPNTIWAQISWMPDTIIFRSTIDGATWTRIMWMTSYPRRSIRYVLIDISAE 360

QY 407 PMLTFGVQPNPVPBPKLGWMDZAMAIDPENSDBMLYGTGATLYATNDLITWDSGGQIHI 466  
 DB 361 PMLTFGVQPNPVPBPKLGWMDZAMAIDPENSDBMLYGTGATLYATNDLITWDSGGQIHI 420

QY 467 APMYKGLBEETAVNDLISPPSGAPLISALGDIAGTTHADVTAVPSTIFTSPPFTGTSDYD 526  
 DB 421 APMYKGLBEETAVNDLISPPSGAPLISALGDIAGTTHADVTAVPSTIFTSPPFTGTSDYD 480

QY 527 AELNPSITIVRAGSFPDPSQPNDRHVAFSTDGKQWFOGSEFGVTTGTVAASADGSRFV 586  
 DB 481 AELNPSITIVRAGSFPDPSQPNDRHVAFSTDGKQWFOGSEFGVTTGTVAASADGSRFV 540

QY 587 WAPGDPQPVVYAVGFGNSMAASQGVPAQAQIRSDRVNPKTFYALSNCTFYRSTDGVTFF 646  
 DB 541 WAPGDPQPVVYAVGFGNSMAASQGVPAQAQIRSDRVNPKTFYALSNCTFYRSTDGVTFF 600

QY 647 QPVAAGLPSSGAVGWFHVPKKGDLMLAASGLYHSTNGSSWSALTYGSSAVNVGFG 706  
 DB 601 QPVAAGLPSSGAVGWFHVPKKGDLMLAASGLYHSTNGSSWSALTYGSSAVNVGFG 660

QY 707 KSAPGSSYPAFVFWGTIGCVTVGAYRSDCGTTWVLINDOHOYGMWGAITGDHANLRV 766  
 DB 661 KSAPGSSYPAFVFWGTIGCVTVGAYRSDCGTTWVLINDOHOYGMWGAITGDHANLRV 720

QY 767 YIGTNGRGIVGDIIGAPSG 786  
 DB 721 YIGTNGRGIVGDIIGAPSG 740

RESULT 6  
 ADD22923  
 ID ADD22923 standard; protein; 740 AA.  
 AC ADD22923;  
 XX 15-JAN-2004 (first entry)  
 DT  
 XX Acidothermus cellulolyticus avicelase AvIII catalytic domain.  
 DE  
 XX enzyme; AvIII; cellulose reduction; agricultural biomass;  
 KM municipal solid waste; glycoside hydrolase; avicelase.  
 XX  
 OS Acidothermus cellulolyticus.  
 XX  
 XX US2003108988-A1.  
 XX  
 PD 12-JUN-2003.  
 XX  
 PF 18-OCT-2002; 2002US-00155400.  
 XX  
 PR 28-JUL-2001; 2001US-00917376.  
 XX  
 XX (DING/) DING S.  
 PA (ADNEY/) ADNEY W S.  
 PA (VINZ/) VINZANT T B.  
 PA (HIMM/) HIMMEL M E.  
 XX  
 PI Ding S, Adney WS, Vinzant TB, Himmel ME;  
 DR WPI; 2003-810853/76.  
 XX



PT New isolated thermal tolerant avicelase polynucleotide useful for  
PT detection of a polynucleotide encoding AvilIII and for reducing cellulose  
PT in a starting material, e.g. municipal solid waste.

PS Claim 16; SEQ ID NO 3; 29pp; English.

CC The invention relates to an isolated polynucleotide molecule encoding a  
CC thermostable AvilIII polypeptide. The polynucleotide is useful for  
CC detection of a polynucleotide encoding AvilIII. The polynucleotide is  
CC useful for reducing cellulose in a starting material which involves  
CC administering to the starting material, e.g. agricultural biomass or  
CC municipal solid waste, a polypeptide molecule of the polynucleotide. The  
CC method further comprises administering a second polypeptide molecule  
CC chosen from the glycoside hydrolase family of proteins. The present  
CC sequence represents the amino acid sequence of Acidothermus  
CC cellulolyticus avicelase AvilIII catalytic domain.

XX Sequence 740 AA;

Query Match 78.6%; Score 4036; DB 7; Length 740;  
Best Local Similarity 100.0%; Pred. No. 3.6e-205;  
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 ATTPTWTSNVAIGGGFVDGIVNEGAPGILYVRTDIGMYRMDANGRWIPLLDWGM 106  
DB 1 ATTPTWTSNVAIGGGFVDGIVNEGAPGILYVRTDIGMYRMDANGRWIPLLDWGM 60  
QY 107 NNMGVGVSVIADPINTNKVMAAVGYTNSWDPNDGAILRSSDQATWQITPLPFXLGG 166  
DB 61 NNMGVGVSVIADPINTNKVMAAVGYTNSWDPNDGAILRSSDQATWQITPLPFXLGG 120  
QY 167 NMPGRGGERLAVDPNNNDILYFGAPSGKGLMRSTDSGATWSQMTNFPDVGYTIANPDT 226  
DB 121 NMPGRGGERLAVDPNNNDILYFGAPSGKGLMRSTDSGATWSQMTNFPDVGYTIANPDT 180  
QY 227 TGYOSDIGVWVWAFDKSSSSLGQASKTIFVGVAADPNNPVFWSRDGAATQAVPGATGF 286  
DB 181 TGYOSDIGVWVWAFDKSSSSLGQASKTIFVGVAADPNNPVFWSRDGAATQAVPGATGF 240  
QY 287 IPHKGVFDPVNHVLYATSNVTGPGYDSSGDVWKFVSTSGTWTRISFPVSTDANDYFGY 346  
DB 241 IPHKGVFDPVNHVLYATSNVTGPGYDSSGDVWKFVSTSGTWTRISFPVSTDANDYFGY 300  
QY 347 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGATWTRIMWTSYPNRLRYVLDISAE 406  
DB 301 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGATWTRIMWTSYPNRLRYVLDISAE 360  
QY 407 PMLTFGVQNPVPVSPKRLGMDDEAMAIDPNSDRMLYGCATLYATMDLTKMDSGGQIHI 466  
DB 361 PMLTFGVQNPVPVSPKRLGMDDEAMAIDPNSDRMLYGCATLYATMDLTKMDSGGQIHI 420  
QY 467 AMVKGLETAIVNDLISPPSGAPLISALGPGFTHADVTAVPSTIPTSVEFTTGTSTVDY 526  
DB 421 AMVKGLETAIVNDLISPPSGAPLISALGPGFTHADVTAVPSTIPTSVEFTTGTSTVDY 480  
QY 527 AELNPSIIVRAGSFDPSSQPNDRHVAFTDGGKMFQGSBPQVTTGGTVAASADGSRFV 586  
DB 481 AELNPSIIVRAGSFDPSSQPNDRHVAFTDGGKMFQGSBPQVTTGGTVAASADGSRFV 540  
QY 587 NARPGDQGPVYVAVGFGNSWMAASQGVPAANAQIRSDRNPRTFYALSNGTFRSTDGVTF 646  
DB 541 NARPGDQGPVYVAVGFGNSWMAASQGVPAANAQIRSDRNPRTFYALSNGTFRSTDGVTF 600  
QY 647 QPVAAGLPSGAGVWVFHAPVKGEGDLMLAASGLYHSTNGSSWSAITGVSSAVNVGFG 706  
DB 601 QPVAAGLPSGAGVWVFHAPVKGEGDLMLAASGLYHSTNGSSWSAITGVSSAVNVGFG 660  
QY 707 KSAFGSSYPAVFVVTIGVYTGAYRSDDCGTTWVLINDDOHQYGNWQOAITGDHANLRV 766  
DB 661 KSAFGSSYPAVFVVTIGVYTGAYRSDDCGTTWVLINDDOHQYGNWQOAITGDHANLRV 720  
QY 767 YIGTNGRGIVYGDIGGAPSG 786  
DB 767 YIGTNGRGIVYGDIGGAPSG 786

DB 721 YIGTNGRGIVYGDIGGAPSG 740

RESULT 7  
AD052315  
ID AD052315 standard; protein; 733 AA.

XX AD052315;

DT 15-JUL-2004 (first entry)

DE A. cellulolyticus avicelase III (AvilIII) GHCA catalytic domain (CD).

KM Thermostable cellulase; avicelase III; AvilIII; cellulose reduction;

KM biomass degradation; ethanol formation; industrial chemical;

XX fabric treatment; catalytic domain; CD.

OS Acidothermus cellulolyticus.

XX US2004038334-A1.

XX 26-FEB-2004.

PF 28-JUL-2001; 2001US-00917376.

PR 28-JUL-2001; 2001US-00917376.

PA (DING/) DING S.

PA (ADNE/) ADNEY W S.

PA (VINZ/) VINZANT T B.

PA (HIMM/) HIMMEL M E.

PI Ding S, Adney WS, Vinzant TB, Himmel ME;

XX WPI; 2004-203224/19.

PT Novel thermostable AvilIII polypeptide of glycoside hydrolase family and

PT isolated from Acidothermus cellulolyticus, useful for degrading treated

PT biomass into simpler forms of carbohydrate.

PS Claim 28; SEQ ID NO 3; 19pp; English.

XX The invention relates to a thermostable cellulase enzyme, avicelase III

CC (AvilIII) and its nucleic acid sequence. AvilIII is useful for reducing

CC cellulose in a starting material. A thermostable AvilIII peptide is useful

CC for degrading treated biomass into simpler forms of carbohydrate, which

CC is used in the formation of ethanol or other industrial chemicals. It is

CC also useful for treating fabrics to remove cellulose-containing stains.

CC The present sequence is Acidothermus cellulolyticus AvilIII protein GHCA

CC catalytic domain (CD).

XX Sequence 733 AA;

Query Match 76.3%; Score 3920.5; DB 8; Length 733;  
Best Local Similarity 99.1%; Pred. No. 4.6e-199;  
Matches 733; Conservative 0; Mismatches 0; Indels 7; Gaps 7;

QY 47 ATTPTWTSNVAIGGGFVDGIVNEGAPGILYVRTDIGMYRMDANGRWIPLLDWGM 106  
DB 1 ATTPTWTSNVAIGGGFVDGIVNEGAPGILYVRTDIGMYRMDANGRWIPLLDWGM 60  
QY 107 NNMGVGVSVIADPINTNKVMAAVGYTNSWDPNDGAILRSSDQATWQITPLPFXLGG 166  
DB 61 NNMGVGVSVIADPINTNKVMAAVGYTNSWDPNDGAILRSSDQATWQITPLPFXLGG 120  
QY 167 NMPGRGGERLAVDPNNNDILYFGAPSGKGLMRSTDSGATWSQMTNFPDVGYTIANPDT 226  
DB 121 NMPGRGGERLAVDPNNNDILYFGAPSGKGLMRSTDSGATWSQMTNFPDVGYTIANPDT 178  
QY 227 TGYOSDIGVWVWAFDKSSSSLGQASKTIFVGVAADPNNPVFWSRDGAATQAVPGATGF 286  
DB 179 TGYOSDIGVWVWAFDKSSSSLGQASKTIFVGVAADPNNPVFWSRDGAATQAVPGATGF 237



QY	287	IPHKGEVDPVNHVLVIATISNTGPPRIDGSSGDWKKSVYSGTWRISPVSTDTANDYRFG	346
Db	238	IPHKG-FPVPVNHVLVIATISNTGPPYGGSSGDWKKSVYSGTWRISPV-STDTANDYFG	295
QY	347	SGLTIDRQHPTIMATQISWMPDITIFSTDSGATWTRIMWTSTVPNLSLRVLDISAE	406
Db	296	SGLTIDRQHPTIMATQISWMPDITIFSTDSGATWTRIMWTSTVPNLSLRVLDISAE	355
QY	407	PMLTFGVQPNPVPSPKLGMDMEAMADFPNSRMLYGCATLYATNDLTKMDGGQIHI	466
Db	356	PMLTFGVQPNPVPSPKLGMDMEAMADFPNSRMLYGCATLYATNDLTKMDGGQIHI	415
QY	467	APMWKGLBETA VNDLISPPSGAPLISALGDGCFTHADYTAVPSTIFTSVPFTGTSVDY	526
Db	416	APMWKGLBETA VNDLISPPSGAPLISALGDG-THADYTAVPSTIFTSVPFTGTSVDY	474
QY	527	AEIAPNSTIIVRAGSPDPSSQPNDRHYAFSTDGKGMVQGSSEPGVTTGCTYAAADSSRV	586
Db	475	AEIAPNSTIIVRAGSPDPSSQPNDRHYAFSTDGKGMVQGSSEPGVTTGCTYAAADSSRV	534
QY	587	WAPGPGQPVVYAVVGFNGSNMAASQGVPAQAQISDRVNPFTFALSNGFVYRSTDGVT	646
Db	535	WAPGPGQPVVYAVVGFNGSNMAASQGVPAQAQISDRVNPFTFALSNGFVYRSTDGVT	594
QY	647	QPVAAGLPSGAGVGVWFHAVPCKEGLDMLAASGLYHSTNGGSSWSAITGVSSAVVFG	706
Db	555	QPVAAGLPSGAGVGVWFHAVPCKEGLDMLAASGLYHSTNGGSSWSAITGVSSAVVFG	654
QY	707	KSAPGSSYPAVFVVGITIGVTA YRSDDCGTTWVLINDDOHQYGNMGOAITGDHANLRV	766
Db	655	KSAPGSSYPAVFVVGITIGVTA YRSDDCGTTWVLINDDOHQYGNMGOAITGDHANLRV	714
QY	767	YIGTNGRGIVGDIGGAPSG 786	
Db	715	YIGTNGRGIVG-IGGAPSG 733	

RESULT 8	
ABB99489	
ID	ABB99489 standard; protein; 940 AA.
AC	
XX	ABB99489;
XX	
DT	03-MAR-2003 (first entry)
XX	
DE	Amino acid sequence of a xyloglucanase enzyme.
XX	
KW	Xyloglucanase; enzyme; family 74; glycosyl hydrolase; cellulosic fiber;
KW	textile scouring; cellulose fiber processing; ratting.
XX	
OS	Jonesia sp.
XX	
PN	WO200277242-A2.
XX	
PD	03-OCT-2002.
XX	
PF	27-MAR-2002; 2002MO-DK000210.
XX	
PR	27-MAR-2001; 2001DK-00000504.
XX	
PA	(NOVO ) NOVOZYMES AS.
XX	
PI	Duffner F, Sjöholm C;
XX	
DR	WPI; 2003-092855/08.
XX	
XX	N-PSDB; ABV76941.
PT	
PT	New xyloglucanase enzyme belonging to family 74 of glycosyl hydrolases,
XX	and endogenous to a bacterium, useful in the textile industry for
XX	improving properties of cellulosic fibers, yarn or fabric.
PS	Claim 5; Page 72-75; 76pp; English.
XX	

The present sequence represents a xyloglucanase enzyme, belonging to family 74 of glycosyl hydrolases. The enzyme is isolated from *Jomsea sp* DSN14140. The enzyme is useful in processes for machine treatment of fabrics. It is also useful in the textile industry for improving the properties of cellulosic fibers, yarn, woven or non-woven fabric, and in textile scouring process step. The xyloglucanase enzyme is also useful in the cellulose fiber processing industry for rattling of fibers such as hemp, jute, flax and linen. It is useful for preventing binding of certain soils to the xyloglucan left on the cellulosic material

**SQ Sequence 940 AA;**

Query Match	42.38;	Score 2174;	DB 6;	Length 940;
-------------	--------	-------------	-------	-------------

Best Local Similarity 44.8%; Pref. No. 1.3e-106;  
Matches 430; Conservative 141; Mismatches 311; Indels 78; Gaps 16;

Dh QY 10 TMRRLVSLAATASFAVAALGVLPILAITASBAHAATQPYTMSNAIGGGFVDGIV 69  
| : : | : ||| | : : : | : : : | : ||| | ||||| |  
5 TVAVATLTAATTAGIAATS--TMPDVTAPPSFSSSSGYSNSNVEIWGCGVPGIV 61

70 PNEGAPGILVVRTDIGMYRWDAANGRWIPLLDWYGMNNMGVGVSTADPINTNKVMA 12

```
Db      62 FNDKDPGLVARTDGCAYRLNDSTGRWIPPLTDHIGDDWSHGILSLATDPVDNRVYL 12
```

130 AVGMYTNSWDGAILLRSSDQATWQITPLPKLGNGMPCRGMRGLAVDPNNINILYF 18

Db 122 AAGTYSMDMPONGALIRSDAGETWEKTMLPFRVGNMPCGROMGERLAVDPNNKKVLYF 18

190 GAPSCKGLMRSTDSGATWSQMTNFPDVGTYIANPDTTG-YQSDIQGVVWAFDPKSSSL 24

```

Db      182 GASSGNGLMSTDYGKTIWGKVTSPFNAGNYVA---DASGAYTQNGGVVWTFEDPTSAKA 23

```

249 GOASKTI FVGVADPNPNVFWSRSDGATWQAVPGAPGTGFI PHKGVFDVNVHVLITATSNIG 30

Db 239 G<sup>1</sup>TT<sup>1</sup>QT<sup>1</sup>TYVGADK<sup>1</sup>NN<sup>1</sup>YRSIDGGATW<sup>1</sup>RVGG<sup>1</sup>PIG<sup>1</sup>FLAQK<sup>1</sup>VL<sup>1</sup>DHK<sup>1</sup>Q<sup>1</sup>LYATSDTG 29

309 GPYDGGSGDYWKFSVTGTRISPVSPSTDANDYFCYSGLTIDRHPNTIMATQISWW 36

Db 299 GPYDGSKGDVWRKLDISSGQWTRISPIPST--SSNSAFQYSGLAIDRKNPDTIMVVSQVSWW 35

QY 369 PDITIFRSTDCGATWIRIMDWISYPNRSRLRVLJISABPWLTFGVQPNBPVPSPLXGMMD 422

Db 358 PDMYYYRSTGRGKTWSPILWELNGSQPRTIQYNHDSGAPWLDIFGNTAKEPEANPKLGWMT 41

429 BAAIDPFNSDRMLYGTGATLYA.TNDLITKNDSSGGQIHIAPMVKLBETAVNDLISPSSGA 48

Db 418 QSPFIDPHNSDRFFYGTGAGIYGGTNLTNNDKGGKVDITVKAQIEETLAQDLAPPGNI 478

489 PLISALGDLGFTHADVTAVPST-IFTSPVFTTGTSVDYAEIINSSIIIVAGSFDSSQPN 54

Db 478 GLYSALADIGGFTHKDISQVFNKYYYKNPHDVTSLDPAESKATVVRAG--SISGET 53

Qy 548 DRHYAFSTDDGKNWFGQSEBPGGVTTCGTVASADSGSFVMAPEBPGQPVVYAVGFGNSWA 600

Db 536 TSNVGVSTDAEETWKPGLTPSGVYKPGSITVSANASSI VWAP -- EGAAPRSTISGSSWS 559

608 ASGGTPANAQLRSDRVNPKTFYALNSNGTFRSTDGCTEQPVA-AGLPSSGAVGVWVHAF 666

594 TWSGPHNAQVASDRVANTLYGEVDGKFYHSINGGASF IASAF IGF FLSG - NVKRF KAV 65

[illegible]

652 PGRQSHLMUAGVSSVSIIGMKRSIDGGNIMIVSGVCEGDAVGFQNAIDGGILEVALLNR 780

713 KTFDPTFEPDECKTAKPILNNHNNOWMGASITGPDVREGRYTGTNRGIIYGD-- 76

789 GABCSBPSPVPSASPT.SPSPSPSSSPSPSPSSSPSSSPSPSP -SPSPSPSP 84

770 -----SSTPPPGDSSGASGNTGATPTPETPDSHNTGC 800

841 SASPQSSSSPSSSSSPPTSSSPVGGGVKVOYKNDSDAPGDNQIKRGLQVNTGS 90



Db 805 STDGETSEBPGTGGSGGTRAPDSGNTSPGNTSSCKRYRSTTDMGSG---FTGAATINTSP 861  
 QY 901 SSVDLSTVTVRY-----WTRD--GGSSTLVYNCWMAAIGCNIRASFGSVNPAIP 949  
 Db 862 SPINQWTLAFTYPSGQTISSVMSATQTLSGRNVVLKNSGM-----NPTIP 906

## RESULT 9

ADD22927  
 ID ADD22927 standard; protein; 726 AA.  
 XX  
 AC ADD22927;  
 XX  
 DT 15-JAN-2004 (first entry)  
 XX  
 DE Aspergillus aculeatus avicelase III catalytic domain.

XX  
 KM enzyme; Aviii; cellulose reduction; agricultural biomass;  
 KM municipal solid waste; glycoside hydrolase; avicelase.  
 XX  
 OS Aspergillus aculeatus.

XX PN US2003108988-A1.

XX PD 12-JUN-2003.

XX PF 18-OCT-2002; 2002US-00155400.

XX PR 28-JUL-2001; 2001US-00917376.

XX PA (DING/) DING S.

XX PA (ADNEY/) ADNEY W S.

XX PA (VINZ/) VINZANT T B.

XX PA (HIMM/) HIMMEL M E.

XX PI Ding S, Adney WS, Vinzant TB, Himmel ME;

XX DR WPI; 2003-810853/76.

XX PT New isolated thermal tolerant avicelase polynucleotide useful for

XX PT detection of a polynucleotide encoding Aviii and for reducing cellulose

XX PS in a starting material, e.g. municipal solid waste.

XX PS Example 2; SEQ ID NO 7; 29pp; English.

XX CC The invention relates to an isolated polynucleotide molecule encoding a  
 CC thermostable Aviii polypeptide. The polynucleotide is useful for  
 CC detection of a polynucleotide encoding Aviii. The polynucleotide is  
 CC useful for reducing cellulose in a starting material which involves  
 CC administering to the starting material, e.g. agricultural biomass or  
 CC municipal solid waste, a polypeptide molecule of the polynucleotide. The  
 CC method further comprises administering a second polypeptide molecule  
 CC chosen from the glycoside hydrolase family of proteins. The present  
 CC sequence represents the amino acid sequence of Aspergillus aculeatus  
 CC avicelase III catalytic domain.

XX SQ Sequence 726 AA;

Query Match 32.7%; Score 1680; DB 7; Length 726;

Best Local Similarity 46.2%; Pred. No. 1.3e-80;

Matches 346; Conservative 113; Mismatches 258; Indels 32; Gaps 17;

QY 47 ATTGPYTWGNVAI-GGGGFTDGIIVNEGAGGIIIVRTDGGMYRMDANGRWIPLIMWG 105

Db 1 AASQATYTKNVVTVTGGGFTGPIVENPSAGVANAARTDGAARLN-SDTWTPLMDMWG 59

QY 106 ---MNNWYGVVSIADPINTKVAWAAGMYTNSMDPNDGAILRSSDQATQITPLPF 162

Db 60 NDTWHDW---GIDALATDPVDTDRVYVAVGMYTNEMDPNNGSILRSIDQDITWETLPLF 116

QY 163 KLGNNPGRGMRGLAVDPNNDNIIYFGADPSGKLWRSTDSGATWSQMTNPPDVGYTIAN 222

Db 117 KVGNNPGRGMRGLAVDPNKNISILYFGARSGHGLWKSITDYGATWSNVTSFTWTGYFOD 176  
 QY 223 PTDTTGYSIDIQVWVAEPKSSSILGQASKITFCVGVADNNPNVFWMSRDGATWQAVGA 282  
 Db 177 SSST--YTSDPVGIAMWTFEDSTSGSSGSAIPRLFVGADGKGVFSSEDCATWAWYSGE 234  
 QY 283 PT-GFIPHKGVPPVNVHVIYATSNNGPFDGSGGVWKPESVTSGTWTRISPVPSDTTAN 341  
 Db 235 PÖYGFIPHKGVLSPEERKTLIYSANGAGPYDGTNGVTHKNITISGVTWTDSP---TSLAS 291  
 QY 342 DYFGSGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGATWTRIMWTSYPNRSLRYVL 401  
 Db 292 TYRYGGSLVDLQVPGLVAVAAALNCWPPDELIIFRSTDSGATWSPIMWNGYPSINYYSY 351  
 QY 402 DISAEPWLTFRGVQPNP-PVPSPKLGMMDEMAALDPEFSDMLYGTATLTATNDLTWDS 460  
 Db 352 DISNAPWIDTSTTDQFPV--RVGMWVEALALDPPDSNMLYGTGLTVYGGHDLTWDS 408  
 QY 461 GGGIHLAPWYKGLERTVNDLISPPSGAPLISALGDLGFTHADVTAVPSTIFPSPVFTT 520  
 Db 409 KHNVTYKSLAVGLEMAVGLITPPGGPALLSAVGDGCFYHSDLDAAFPQATHTPTTGT 468  
 QY 521 GTSVDYALNPSTIVRAGSFDPSSOPNDRHVAFTDGAKNWFOGSEPGVTTGGTVAASA 580  
 Db 469 TNGIDYAGNKPSTIVRAGSAD--DYPT--LALSSNGSTWYADVAASTGTGAVALSA 523  
 QY 581 DGRFVWAPGDPQOPVYVAVFGNSMAASGCPANAOIRDRVNPKTFFYLSNCTFFRST 640  
 Db 524 DGDVTLLMSTSGALVSKSQG--TLTAAVSLSPGAVIADKSDNTVFYGGSAIAIYVSK 580  
 QY 641 DGVTPQVPAAGLPSGAGVGMFHAIVKRGDMLLAASGLYHSTNGGSSWSAI--TVSS 699  
 Db 581 NTATSPFTKTVS-LGSSITTVNAI-RAHPSIAGDVAASTDKGLMSTIDGSTTQIGSGVTA 638  
 QY 700 AVNWGFGKSAFGSSYPAPVFWGTIGVTGAYRSDCCFTWVLINDDOHYGNMGOA--IT 757  
 Db 639 GMSRFGKASSTGSGYVYIGFTTIDGAAGLFKESDADATNNQVVISDAHGFSGS-GSANNVN 697  
 QY 758 GDHANLIRYVYITNGRGIYVGDIGAPSG 786  
 Db 698 GDLQTYGRVFRGHERPGLLRQSGRBPAG 726

## RESULT 10

ADD42055  
 ID ADD42055 standard; protein; 838 AA.

XX AC ADD42055;

XX DT 15-JAN-2004 (first entry)

XX DE Trichoderma reesei Family 74 xyloglucanase.

XX KW Trichoderma reesei; Family 74 xyloglucanase; beta-1,4-glycosidic linkage;

XX KW xyloglucan hydrolysis; biomass degradation; cellulose; hemicellulose;

XX KW ethanol production; detergent composition; fabric treatment;

XX KW textile treatment; enzyme.

XX OS Hypocrea jecorina.

XX FN WO2003089598-A2.

XX PD 30-OCT-2003.

XX PF 17-APR-2003; 2003WO-US011831.

XX PR 19-APR-2002; 2002US-0373987P.

XX PA (NOVO ) NOVOZYMES BIOTECH INC.

XX PI Michael R, Zaretsky E, Haas J;

XX WPI; 2003-845528/78.



DR N-PSDB; ADD42054, ADD42060.

XX New polypeptides having Family 74 xyloglucanase activity, and encoding  
PT nucleic acid molecules, useful for degrading cellulose- and hemicellulose  
PT -containing biomasses to ethanol or as a detergent.

PS Claim 8; SEQ ID NO 2; 96pp; English.

XX The invention relates to a Family 74 xyloglucanase (Add42055) from the  
CC fungus *Trichoderma reesei* (Hypocrea jecorina), and nucleic acids encoding  
CC it (e.g., ADD42054). The enzyme hydrolyses the beta-1,4-glycosidic  
CC linkages in the backbone of xyloglucan to xyloglucan oligosaccharides.  
CC The invention also relates polypeptide sequences at least 70% identical  
CC to the enzyme, expression vectors and host cells comprising a nucleic  
CC acid of the invention, the recombinant production of the enzyme, and  
CC mutant enzymes and the nucleic acids encoding them. The xyloglucanase of  
CC the invention can be used in the degradation of cellulose- and  
CC hemicellulose-containing biomass to produce ethanol. It can also be used  
CC in a detergent composition for treating fabric during a machine washing  
CC cycle. The present sequence represents the *Trichoderma reesei* Family 74  
CC xyloglucanase.

XX Sequence 838 AA;

Query Match 31.7%; Score 1625.5; DB 7; Length 838;  
Best Local Similarity 41.7%; Pred. No. 1.1e-77;  
Matches 350; Conservative 131; Mismatches 296; Indels 63; Gaps 23;

25 SFVAALGLVLAITASPAHAATTOPTYSNVAI-GGGFVDGIVENEGAGILLYRTD 83

4 SRLALVLGVN-----IPAHAA-----FSKNVNLGGGGGVPEIIRHPKTKGVAVARTD 53

84 IGGMYRDAAANGEMIPLLDNV---GMNNGYNGVSIADDPINTKVMVAAGMYTNSWD 139

54 IGLYRLN-ADDSWTAATDGIADNAGWHNN---GIDVVALDPODDQRYKVAAGMYTNSWD 109

140 PNDGALIRSSDQATQITPLPFLGNNMGRGGERLADPNNDNLLYFEGAPSGKGLMR 199

110 PSNGALIRSSDRGATWSFTNLPRKVGNNMGRGGERLADVPANSNIIYFARSNGGLMK 169

200 STDSGATWSQMTNPPDVGTIANPTDTGYOSDIQGVVVAVFDKSSSLGQASRTIIFGV 259

170 STDGVTFRKVSFTATGTIIPDSDSNGNSDKGLMWTTFDSTSTGGATSRIFVGT 229

260 ADP-ANPWFMSRDGATWQAVPGAFTGFIPIHKGVFDPVNHVLYATNSGTGYDSSGDV 318

230 ADNITASVYVSTNAGSTWASAVPGQPKYFPHKALQPAEKALYLYTSGTGPTGLGSV 289

319 WKSSTVSGTWTTRISPVSTDTANDYFGYSGLTIDRQHPNTIMATQISWPDITIFRST 378

290 WRVYDIAGTWDITPVSGSDI--YFGPGGLGLDLOKRGILVVASLMSWPDADLFRSTD 346

379 GGAITTRIMDTSYPNRSLRYVLDISABPMLTFG-VQNPVPVPS-----PKGMDEAMAI 433

347 SGTWSTWIMWASIPETIYYISITPKAPWKKNFIDVTSSPSDGLIKRGMHIESIEI 406

434 DPFNSDMLYGTGATLYATNDLTKMDSGQIHIAPMVKGLSEETVNDLISPPSGAPLISA 493

407 DPFDSNMWLYGTGATLYATNDLTKMDSGQIHIAPMVKGLSEETVNDLISPPSGAPLISA 466

494 LGDLGFTTHA---DVTAVPSTIPTSPTFTGTSVDYALBNLSIIVRAGSPSSSQPNDRH 550

467 VGDNGFTFASRNDLGTSPQVWATPTWATSTVDYAGNSVYSVVRVNTAGTQO----- 521

551 VASTDGGKMWQSSBEGVTTGTGTVAASADGRFWAHPGPQGVYAVAFGNSWASQ 610

522 VAISSDGATWSIDYAADTSMNGGTIVYASADGDTILMSTASSG---VQRSGFQSSFASVS 578

611 GVPANAOIRSDRVNPKTFYALNSGTFRSTGGVTFQVVAAGLPSSGAVGVN--FHAVPG 668

579 SLPRGAVIASDKTNSVFIYAGSGSTFYYSKDTGSSF---TRG-PKLSAGTIIRDAIAPT 634

669 KEGDLMLAASGLYHSTNGSSWSAI--TGVSANVNVGFGKAPGSSYPVAVVGTIGVVT 727

Db 635 TAGTLVSTDVGIIFRSTDSGTFPQGVSTALNTYQIALGVGS-GSNW-NLYAFGT--GPS 690

Qy 728 GA--YRSDDCGTTVVLINDDOHQYGNWGOAITGPHANIRRYIGTNGRGIVY--GDIGGA 783

Db 691 GARLYASDGSASMTDQIGSQSPSISSTVYAGSGSTAGQYVYVTTNRRGIFVYQGYVGG 750

Qy 784 PSGSPSPSVSPASPSISPPSPSSPSPPSPSSPSPPSPSPSPSPSPSPSPSPSPSAS 843

Db 751 TCGT-----SSSTKQSSSTSSASSSTTLKSSVSTTRASTVSSRTSSAAGPTGS 801

RESULT 11  
ADH51581  
ID ADH51581 standard; protein; 818 AA.

AC ADH51581;

DT 25-MAR-2004 (first entry)

XX Trichoderma reesei EGV1 protein sequence related to ethanol production.

XX EGV1 protein; endoglucanase; ethanol production; biomass composition;

XX fermentation; cellulohydrolase; detergent production; softening agent;

XX cotton fabric; wood pulp degradation; sugar; enzyme.

XX Hypocrea jecorina.

XX Key Location/Qualifiers

FT Misc-difference 1..818  
FT /label= OTHER  
FT /note= "OTHER= All Xaa's given in this sequence are  
FT unknown amino acids, given as J in the specification.  
FT They should probably be Ile."

XX US2003113732-A1.

XX 19-JUN-2003.

XX 18-DEC-2001; 2001US-00026994.

XX 18-DEC-2001; 2001US-00026994.

XX (DINN)/ DINN-COLEMAN N.

XX (GOED)/ GOEDGEBOUR F.

XX (WARD)/ WARD M.

XX (YAOJ)/ YAO J.

XX Dunn-Coleman N, Goedgebuer F, Ward M, Yao J;

XX WPI; 2004-106460/11.

XX N-PSDB; ADH51580, ADH51583.

XX New substantially purified EGV1 polypeptide with the biological activity  
XX of endoglucanase, useful for degrading biomass to ethanol.

XX Claim 18; SEQ ID NO 2; 26pp; English.

XX This invention relates to a novel purified EGV1 protein with the  
XX biological activity of an endoglucanase, comprising an amino acid  
XX sequence chosen from a fully defined sequence as given in specification.  
XX The protein is useful for producing ethanol, which involves contacting a  
XX biomass composition with an enzymatic composition containing the protein  
XX of the invention to yield a sugar solution, adding to the sugar solution  
XX a fermentative microorganism, and culturing the fermentative  
XX microorganism under conditions sufficient to produce ethanol, where the  
XX biomass composition may be optionally pretreated. The method further  
XX involves the addition of at least one endoglucanase or cellulohydrolase.  
XX The pretreatment is with a dilute acid. The protein of the invention is  
XX also useful for the production of a detergent composition which may be  
XX used as softening agent and for improving the feel of cotton fabrics and  
XX for degrading wood pulp into sugars. The present sequence is that of the  
XX T reesei EGV1 protein of the invention.



XX Sequence 818 AA; 30.9%; Score 1588; DB 8; Length 818;  
SQ Query Match Best Local Similarity 41.6%; Pred. No. 1.1e-75;  
Matches 338; Conservative 126; Mismatches 295; Indels 54; Gaps 21;

QY 52 YTWNVAI GGGGFDGIVFNAGPGLIYVRDIDGMYRMDANGWIPILDMV---GW 106  
DB 2 FSWKNVKLGCGGFPVGLIFHPKTKGVAARFDIGLRYLNL-ADDSMTAVTDGIADNAGW 60  
QY 107 NNMGNVAVSIADPNTNKVAAVGMATNSWMDPNSGALIRSSDOGATWQITPLPKLGG 166  
DB 61 HNW--GIDVALDPODDQKVAAGVATNSWDPNSGALIRSSDOGATWQITPLPKLGG 117  
QY 167 NMPGRGMBRLAVDPNNNDNIIYFGAPSGKGLMRSTDSGATWSQMTNPDVGIYIANPTDT 226  
DB 118 NMPGRGMBRLAVDPNNNSIIYFGAPSGKGLMRSTDSGATWSQMTNPDVGIYIANPTDT 177  
QY 227 TGYSDIDQGVAVAPDKSSSLGQASKTIFVGVADP-NMPVMSRDOGATWQAVPGAPTG 285  
DB 178 NGYNSDKQIGMWTPESTSTTGATSRIFVGTADNITASVYVSTAGSVAAPGQPGK 237  
QY 286 FIPHGVPDPVNHVYIATSNITGAPYDSSGDVWKFVSITSGTTRISPVSTDPADYFG 345  
DB 238 YFPHAKIQPAEKALYLYSDGTGPYDGLGSWMRYDIAGIKDXITPVSGDL---YFG 294  
QY 346 YSGLTIDRQHPNTIMVATOISWMPDTIFRSTDGATWTRIMWTSYPNKSLRYVLDISA 405  
DB 225 FCGLGIDLOKPGTLVVAASLNSWMPDQPLFRSTDSGTTWSITWMASTPTETYYISISTPK 354  
QY 406 EFWLTFG-VQNPVPVS---PLGWMDEMAIDPNSDMLYGTGATLYATNDLTKWDS 460  
DB 355 APMIKNNFIDVISESPSDGLIKRIGWMIISLEIDPDSNMLYGTGATLYGCHDLITWMDT 414  
QY 461 GGOIHAPVWKGLEFAVNDLISPPSGAPLISALGULGFTHA---DVTAPSTIPTSVP 517  
DB 415 RHNVSIQSLADGIEEFSVODLASAPGSEILAAVANGFTFASRNDLGTSPQVWMAPT 474  
QY 518 FTTGTSVDYAEILNPSTIIVRAGSFPDSSQPNDRHVAESTDGAQKMPFGSEFGVTTGTV 577  
DB 475 WATISVVDYAGNSVKSVRVGN-----TAGTQVALISSDGAQKMSIDYADTSMNGTVA 528  
QY 578 ASADGRFVWAPDPCQPVVYAVGVGNSMAASQGVPAQAQIRSDRVNPKTFYALSNCTFY 637  
DB 529 YSADGDTILMSTASG---VQSRQFQGSFASVSLIPAGAVIASDKKTNSEFYAGSGSTFY 585  
QY 638 RSTDGVTQPVAAAPSPSGAVGV--FHAVPQKEGMLMASSGLYHSTNGSSWAI- 694  
DB 586 VKRDTGSSF---TRG-PKIGSAGTIPDIAHPTTAGTLVYSTDVGIYRSTDSGTTFGQVS 641  
QY 695 TGVSASAVNVGFGKAPGSSYPAVFVVGITGATGA--YRSDDCGTTVWLINDDOHQGNW 752  
DB 642 TALITNTYQIALGVGS-GSNW-NLYAFGT--GPSGARLYASGDSGASTDIDGSGSFGSID 697  
QY 753 GOAITSIDHANLRRVYIGTNGRGIVY-GDIGAPSGSPSPSVSPASPSLSPSPSSSP 810  
DB 698 STKVASGSGTACGVVYGTNGRGVYAGTGVGGGTGT-----SSSTQSSSSSTSSA 748  
QY 811 SPSPSSSPSSSP 843  
DB 749 SSSSTLRSSVSVSTTRASTVTTSRTSSAAGPTGS 781

RESULT 12  
ADH19118  
ID ADH19118 standard; protein; 812 AA.  
XX  
AC ADH19118;  
XX  
DT 11-MAR-2004 (first entry)  
XX  
DE Geotrichum sp M128 xyloglucan oligosacch-degrading enzyme pre protein.

XX  
KM xyloglucan oligosaccharide degradation; enzyme; genetic engineering;  
KW precursor.  
XX  
OS Geotrichum sp. M128.  
XX  
PN EPJ350844-A2.  
XX  
PD 08-OCT-2003.  
XX  
PF 25-MAR-2003; 2003EP-00251866.  
XX  
PR 25-MAR-2002; 2002JP-00083433.  
XX  
PA (NRAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.  
XX  
PI Yaoli K, Mitsunishi Y;  
XX  
PI WPI: 2004-100948/11.  
XX  
DR N-PSDB; ADH19117.  
XX  
PT Novel xyloglucan oligosaccharide-degrading enzyme with a different  
XX degradation mechanism from known enzymes.  
XX  
PS Claim 5; SEQ ID NO 12; 39pp; English.  
XX  
CC The invention relates to a novel xyloglucan oligosaccharide-degrading  
XX enzyme with a different degradation mechanism from known enzymes. The  
XX primary structure and polynucleotide structure of the xyloglucan  
XX oligosaccharide-degrading enzyme provided by the present invention allows  
XX high-purity polypeptides having a xyloglucan oligosaccharide-degradation  
XX activity to be prepared at a low cost through a genetic engineering  
XX process. The current sequence is that of the Geotrichum sp. M128  
XX xyloglucan oligosaccharide-degrading enzyme precursor protein of the  
XX invention.

XX  
SQ Sequence 812 AA; 21.9%; Score 1125; DB 8; Length 812;  
Query Match Best Local Similarity 34.5%; Pred. No. 3.2e-51;  
Matches 285; Conservative 129; Mismatches 291; Indels 120; Gaps 31;

QY 31 ALGVLPFAITASPAHAATTCQYTWNSVAIGGGGVDVFNENAGPGLIYVRDIDGMYRW 90  
DB 6 SLGALFALSLASIAVAKETFEKNVAIGGGYITGIVAHPKTKDLIYARTIDGAYRW 65  
QY 91 DAANGRWIPLLDWGMNMGYNGVVSIAADPINTNKVAAVAMWT-NSWDPNDGALRSS 149  
DB 66 DAGTSKTIPLNDFIEADDMNMGTESTIALDPNNDRLYLAQRVVGDEM---AIFYSE 121  
QY 150 DQCATWQITPLPFKLGMPGRGMBRLAVDPNNNDNIIYFGAPSGKGLMRSTDSGATWSQ 209  
DB 122 DRQGSFTIYESPFPMGANDMGRNNGERLAVNPFSNENVMGTRT-EGIMKSSDRAKTWTN 180  
QY 210 MTNPDVGTIYIANPTDTTGYSDIQGVVAVAFDSSSSSLGQASKTIFGVADPNNPVFWS 269  
DB 181 VTSIPDAF-----TNGIGYTS-----VLFDP-----BRNGTISATAPQG-MVY 220  
QY 270 RDGATWQAVPGATGFI-----PHKGVFDPVNHVLYIATSNITGAPY 311  
DB 221 HDGVSMEPVAGQSSSWLNRTTGAFPDKKASIAQPHKVALTP--NFLVYTYADYPPFW 278  
QY 312 DGSSGDVWKFVSITSGTWTRI-----SPVSTDTANDYFGYSGLTIDRQHPNTIMVATO 364  
DB 279 GVTGGEVWRQRTSGAMNDIIPRVGNSSPAPVYNTQTFPAGFCGASVATVAPNRLVIT- 337  
QY 365 ISWMPDTI---IFRSTDGATWTRIMWTS-----YNNRSIRYVLDISABPWLFG 412  
DB 338 LDRDPGALDSIYLTSDAGATWQVTOQLSSPSNIEGNGKHPNARY-KDGTVPWLPFN 396  
QY 413 VQP-----NRPVPSP---KLGWDEMAAIDPNSDRMLYGTGATLYATNDLTK---WDS 460  
DB 397 NGPQWGCYGAHGHTGTLTKFGWMSAVLIDFPNPHLMYGTGATIMWIDTILSRVAKDM-- 454







XX Synthetic.  
 OS Geotrichum sp. M128.  
 XX EP1350844-A2.  
 XX PD 08-OCT-2003.  
 XX PF 25-MAR-2003; 2003EP-00251866.  
 XX PR 25-MAR-2002; 2002JP-00083433.  
 XX (NAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.  
 XX Yaori K, Mitsunishi Y;  
 XX WPI; 2004-100948/11.  
 DR N-PSDB; ADH19123.  
 XX  
 PT Novel xyloglucan oligosaccharide-degrading enzyme with a different  
 PT degradation mechanism from known enzymes.  
 XX  
 PS Claim 6; SEQ ID NO 18; 39pp; English.  
 CC The invention relates to a novel xyloglucan oligosaccharide-degrading  
 CC enzyme with a different degradation mechanism from known enzymes. The  
 CC primary structure and polynucleotide structure of the xyloglucan  
 CC oligosaccharide-degrading enzyme provided by the present invention allows  
 CC high-purity polypeptides having a xyloglucan oligosaccharide-degradation  
 CC activity to be prepared at a low cost through a genetic engineering  
 CC process. The current sequence is that of the Geotrichum sp. M128  
 CC xyloglucan oligosaccharide-degrading enzyme-derived protein of the  
 CC invention which comprises the wild-type enzyme sequence plus a C-terminal  
 CC region including a Histidine tag originating from vector pET29a(+).  
 XX  
 SQ Sequence 826 AA;  
 Query Match 21.7%; Score 1114; DB 8; Length 826;  
 Best Local Similarity 35.0%; Pred. No. 1.2e-50;  
 Matches 281; Conservative 123; Mismatches 280; Indels 120; Gaps 31;  
 QY 52 YWMSNVALIGGGGVDGIVNEGAPGLVYRTDTCGMVRRDANAGRMIPILDMVGMNNWGY 111  
 DB 5 YEPKVNALIGGGGTTGIVAHPTKXDLVYARTDTCGAYRWDAGTSKMLPLMDFTLEADQMT 64  
 QY 112 NGVSIADPINTNKVAAVGMYT-NSWPDNDGAILRSSDQATWQITPLPFLKLGMPG 170  
 DB 65 MGTESIALDPNPNDRXYLAQGRVGDW---AAFVSEDRGQSFTIYESPFMGANDMG 120  
 QY 171 RQNGERLAVDPNNDNLVFGAPSGKGLMRSTSGATWSQNTNFPDVGTYIANPTDTTGYQ 230  
 DB 121 RNNGERLAVNPFSNVEVMWGTRT-EGIKSSDRAKTWNTVTSIPDAF-----TNGIGYT 173  
 QY 231 SDIQGVVWVAFPKSSSLQASKTIFGVADPNNPVFWSDGATQAVGAPRGFI--- 287  
 DB 174 S-----VIFDP-----ERNGTIVASATAPQG-MYVTHDGVSEVPVAGQBSWLNRT 219  
 QY 288 -----PHKGVFDPVNHVLYATNTGPGPYDSSGDVWKFVSTGWTIRI- 331  
 DB 220 TGAFFPKKPASTAPQPMKVALTP--NFLVYTVADYFGQPMKVTVTGKWKQRKRTSGAMDIT 277  
 QY 332 -----SPVPSDTDTANDYFGYSGLTIDRQHPNTIMATQISWMPDTI---IFRSTGQAT 382  
 DB 278 PRVGNSSPAPYNNQTFPAGGFCGLSVDATNPMLVYIT-LDRDPGALDLSIYSTDAGAT 336  
 QY 383 WTRIMWTS-----YPNRSLRYVLDISAEPLTFGVQP-----NPVVPSP---KLG 425  
 DB 337 WXDVTQSSPSNLEGNMGWGHPTNARFY-KDGTVPVWIDFPNNGPMGWGAGHGPGLTKFG 395  
 QY 426 WMDENAIIDPFNSDRMLYGTGATLVATNDLTK-----WDSGQGHIAF-----MYGKLBETA 477  
 DB 396 WMSAVLIDPFNPEHMLYGTGATIMATDITLSRVEKDW-----APSWTLOIDIGIEBNA 447

QY 478 VNDLISPPSGAPLISALGDIGFTHADVTAVPSTIIFTSPVFTTGSTVDAELNPSIIVRA 537  
 DB 448 ILSLRSPKSGAALLSGIDISGMKHDLTR-PQMFAPQPSNDSIDDAENFNNVYRA 506  
 QY 538 GSPDPSQPNDRHVAFSTDGKRW--FQSEPGQVTT--GGTYAASADGSRFTVAP--G 590  
 DB 507 GSSGHEVDSACARGAYATDGGDAWTFPTCPGKNASHYQSGSTLAVDASGQIYWSTRLD 566  
 QY 591 DPGQPVVYAVGFQNSMAASQVPA-----NAQIRSDRVNPKFETALNSGFTFRSTQGV 644  
 DB 567 EQASGPMYSHDYGTWS-----VPAGDIKAOTANVLSKVDDGTFYATDGGKFEFVSTDGK 622  
 QY 645 TFGPVAAGLPSSGAVGVWFAVPEKEDLWL-AASGLYHSTNGSSWSAI-TEVSSAVN 702  
 DB 623 SYAAKAGLVY--GTSLMPAVNPVAGDVWVPVEGGLFHTSDGASFTRVGTATNLVS 680  
 QY 703 VGFGRK---APGSSYPVAVVGT--IGVTGAYRSDCGTTWVLINDDOHQYGNWQAI 756  
 DB 681 VGAPKSKSDGKKASAPSAVFWMGTDRKPSDIGLYRSDNGSITWRVNDQEHNVSG-PTMI 739  
 QY 757 TGDHANLRVYIGTNGRGIVYGD 780  
 DB 740 EADPKVYGRVYLTGNGRGIVYADL 763  
 RESULT 15  
 ADR90296 standard; protein; 776 AA.  
 ID ADR90296 standard; protein; 776 AA.  
 AC ADR90296;  
 XX  
 DT 16-DEC-2004 (first entry)  
 XX  
 DE Geotrichum sp. M128 xyloglucanendohydrolase precursor protein.  
 XX  
 KM endo-type xyloglucanendohydrolase;  
 KM xyloglucan-specific endo-b-1,4-glucanase; glucose; xylose; xyloglucan;  
 KM plant cell differentiation; precursor; enzyme.  
 XX  
 OS Geotrichum sp. M128.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..20  
 FT Protein label= Signal peptide  
 FT note= "Geotrichum sp. M128 xyloglucanendohydrolase  
 FT mature protein"  
 XX  
 PN JP2004261037-A.  
 XX  
 XX 24-SEP-2004.  
 PD  
 XX  
 PF 28-FEB-2003; 2003JP-00053286.  
 XX  
 PR 28-FEB-2003; 2003JP-00053286.  
 XX  
 PA (DOKU-) DOKURITSU GYOSEI HOJIN SANGYO GIJUTSU SO.  
 DR WPI; 2004-665466/65.  
 DR N-PSDB; ADR90295.  
 XX  
 PT Novel xyloglucan hydrolase derived from Geotrichum genus, having ability  
 PT to hydrolyze beta-1, 4-glycoside linkage of glucose residue in  
 PT xyloglucan, useful for elucidating structure or function of xyloglucan.  
 XX  
 PS Claim 5; SEQ ID NO 13; 86pp; Japanese.  
 XX  
 CC The invention relates to a novel xyloglucanendohydrolase (xyloglucan-  
 CC specific endo-b-1,4-glucanase) enzyme derived from the Geotrichum genus  
 CC and having the ability to hydrolyse the beta-1,4-glycosidic linkage of a  
 CC glucose residue, but not that of a xylose residue contained in a  
 CC xyloglucan. The polypeptide of the invention may be useful for  
 CC elucidating the structure or function of a xyloglucan having an important



role in the differentiation of a plant cell. The current sequence is that of the Geotrichum sp. M128 xyloglucanendohydrolase precursor protein of the invention.

Sequence 776 AA;

Query Match 19.3%; Score 993; DB 8; Length 776;

Best Local Similarity 32.4%; Pred. No. 2,9e-44; Matches 273; Conservative 127; Mismatches 289; Indels 154; Gaps 34;

```
QY 11 MRSRLVSLAATAAFVAALGVLPATASPAHAATQPYTWSNVAIGGGFVDGIVF 70
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4 LRTFGLALVCAVGSWAVTABL-----KPYTISGGGFIISGLVA 41

QY 71 NEGAPGLVPRDIDGATWRMDAANGWIPILDV-----GMNMGYNGVSIADPIIN 125
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 42 HPTKDLIYARDIDGTYRMNAKMEWEPITDFIINNALAGNANLGTESIALDPHPND 101

QY 126 KVMVAVMWYNSMDPNDGAILRSSDQATWQITPLPKLGNNMFGMGERLAVDPNNDN 185
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 102 RLYLAQGDIV-QMDP-MAAFVLSDDRGKTFKQYRSPVPMGANDMGRNGERLAVNPHWT 159

QY 186 ILYPGAPSGKGLMRSTSGATWSQMTNFPDVGTIYIANPTDTGYQSDIQGVWVAFDKSS 245
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 160 ELMPGSRIT-QGLMRSTDRAGTWSRMNQLPDSSIT-----GI-----GIISVIFDPKN 205

QY 246 SSLGQASKTIFVGVADPNNPVPFWSRDGATWQAVPGAF-----G 285
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 206 VGTAYVAASHAVGLW-----VTW--DGANMSQVGGQPTQMSDWTKSIVAASGTALQSSG 258

QY 286 FIPHKVFPVNVHVIYIATSNQGPYDGSQDVWKPSVTSGTWRISFVPSTDFANDY-- 343
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 259 PLPIKIALGK-NGRLITYSDAPGWPGLYGEVWSYDPTGNMKHIT--PSRGGANTYPA 315

QY 344 -----FGYGLTIIDROHPNTIMATQISWMPDTIIFRSTDGATWTRIMDTSYPN 394
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 316 PTGNKKVVPGGNNGISVG--NGDTVVYSTIDANGEDS-VYLSRDAGNSWKDLGKLTTPAG 372

QY 395 RSLRYVLIDISAE-----PMLTFGVQPNPVPSPKLGMDDEAMALDPFNSDRMLYGTGA 447
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 373 AGNSQKESDAXLRNGTPLPLWLSFQNRGSGIV--GFGMWLAAILLDPF-SDRLLYGTGA 428

QY 448 TLVATDILTKMDSGGQIHIAFM-----VKGLETAVNDLISPPSG-APLISALDGLGFTH 502
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 429 VFWATDAVGRADS---NQAPSWYINTEGIEETAILVLSKPPAGPAHLFSGMYDLGGWRH 484

QY 503 ADVTAVPSTIFTSVPFTTGSVDYAEINPGLIIVRAGSFD--PSSQPNDRHVAFSFDGKN 560
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 485 DDF-SVPQPMYSKPTSTSGLDPAAGRAANVLARVGRNDHPDAGVAGCTOGAYTTNSGDS 543

QY 561 W--FQGSERQ-GVTTGTVAASADGSRFVAPGD-PGQPVVYAVFGNSWMAASQGVAPNA 616
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 544 WTLFQTCVPSLEVGNGSTIAVGADGKTFVWSPSKADGKGPYTSDDYCKWTAPSGLSKQT 603

QY 617 Q-IRSDRVNPKTFYALNSGTFYRSTDCGVTFPQVPAGLPSSGAVGMFAVPG---KEG 671
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 604 TGIADRVQANTFYVVEGDFVFVSTDGSKSYTKKNGLPCCWT---YTGTPVTSNLRAG 659

QY 672 DLMLAASS-GLYHSTNGSSMSALITGVSSAVNVGFGKAPSSYPAYFVVGT----- 722
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 660 ELWVSVKGVGIYHSTDFGNTFTALAGSSSLN-----PAVFSIGAPQTPNATE 707

QY 723 ---IGV-----TGAVRSDCGTTWVLINDDQHQYGNWCAITGDHANLRRVYIGTNGR 773
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 708 TLFKNGIPSAQPEGLYMSTDNGMLTRINDDAINYG-ATVISGDPRIYGRVYIGMNGR 766

QY 774 GIV 776
   : :
Db 767 GII 769
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 07:48:13 ; Search time 26.555 Seconds  
(without alignments)  
3467.504 Million cell updates/sec

Title: US-09-917-376-1

Perfect score: 5135

Sequence: 1 MDRSENIRLTKMSRRLVSL.....RASFGSVNPAFTADTYLQX 957

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR 79: \*  
2: PIR1: \*  
3: PIR2: \*  
4: PIR3: \*  
5: PIR4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2436	47.4	890	2 T35237	probable secreted
2	2037	39.7	839	2 D97013	probable secreted
3	1740	33.9	856	2 T00349	Avicelase III - As
4	1161	22.6	707	2 F72393	hypothetical prote
5	431.5	8.4	1779	2 T31085	xylinase - Caldice
6	398.5	7.8	1711	2 T31337	1,4-beta-glucanase
7	397.5	7.7	1742	2 T17120	cellulase (EC 3.2.
8	372.5	7.3	1743	2 S50755	hypothetical prote
9	346.5	6.7	901	2 A49227	salivase - Actino
10	340.5	6.6	1331	2 A48954	mannan endo-1,4-be
11	308.5	6.0	611	2 S76211	hypothetical prote
12	301.5	5.9	1749	2 S75138	hypothetical prote
13	293	5.7	279	2 T10361	hypothetical prote
14	291	5.7	351	2 S50754	hypothetical prote
15	290.5	5.7	2232	2 T34434	hypothetical prote
16	290	5.6	1039	2 S02711	cellulase (EC 3.2.
17	288.5	5.6	1032	2 T34433	hypothetical prote
18	283.5	5.5	474	2 S15921	protein TPX-VT3 -
19	262	5.1	2468	2 A83412	hypothetical prote
20	261.5	5.1	1854	2 S36859	c1pA protein - Clo
21	254.5	5.0	913	2 S20582	exo-alpha-sialidas
22	246	4.8	915	2 A43802	cellulase (EC 3.2.
23	245	4.8	879	2 A47704	endoglucanase I (B
24	244	4.8	1664	2 T18262	S-layer protein -
25	243.5	4.7	552	2 T08148	proline-rich myos
26	243	4.7	288	2 T17737	proline-rich prote
27	235.5	4.6	2124	2 A28452	proteoglycan core
28	229.5	4.6	496	2 T17908	proline/lysine-ric
29	227.5	4.4	2271	2 F90073	hypothetical prote

30	222.5	4.3	360	2 S12850	protein TPX - Ther
31	222.5	4.3	3570	2 T45025	mucin MUC5B, trach
32	221	4.3	2132	1 A55182	aggregan precursor
33	219	4.3	4776	2 E95206	cell wall surface
34	218.5	4.3	3164	1 WMBEH6	U36 protein - hum
35	217	4.2	241	2 AC2284	hypothetical prote
36	216	4.2	383	2 T46707	proteophosphoglyc
37	216	4.2	449	2 S16748	proline-rich prote
38	215.5	4.2	13055	2 T16580	hypothetical prote
39	215	4.2	339	2 T17636	proline-rich prote
40	215	4.2	13288	2 T03099	mucin, submaxillar
41	214	4.2	300	2 T02220	hydroxyproline-ric
42	214	4.2	532	2 B35621	spore germination
43	211.5	4.1	1275	2 T13369	hypothetical prote
44	211	4.1	602	2 AD2067	hypothetical prote
45	210	4.1	549	2 T17525	proline-rich prote

#### ALIGNMENTS

##### RESULT 1

T35237

probable secreted cellulase - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 09-Jul-2004

C:Accession: T35237

A:Reference number: 221572

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Residues: 1-890 <SEE>

A:Cross-references: UNIPROT:086727; EMBL:AL031515; PIDD:CAA20642.1; GSPDB:GN00070; SCOE

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOE:DB:SC5C7.30C

Query Match	Query	Match	Score	DB 2;	Length	890;
Best Local Similarity	48.6%;	Pred. No. 2.3e-101;				
Matches	475;	Conservative 129;	Mismatches 248;	Indels 126;	Gaps 19;	
QY	11 MRSRLVSLAATSAFAVAALGVLPITATSPA--HAATQPTWTSNVAIGGGFVDCI 68					
DB	1 MERTRLTVLALAAGLA--GSPPAASAAPRAVAADSYTKAKRIQGGFVDCI 57					
QY	69 VFNKGAPILYVTDIGMYMDAANGRWIPLLDWGNMNGVNSIAADPINTKRW 128					
DB	58 VFNREKDLAARTDIGAYRQESHTWTPLDHDVGMDDGHTGVVALASDAVDPDRY 117					
QY	129 AAVGMTNSWPDNCAIIRSSDQATWQITPLPFLGGMPPRGKGERLAADPNNDILY 188					
DB	118 AAVGTYTNDPPTNGAVLRASDRGASWEKADLPFLGGMPPRGKGERLAADPNNDILY 177					
QY	169 FGABSGKLMSTDSGATWQMTNFPDVGTYIANTPTTGYOSDIQGVVVAFDKSS--SS 247					
DB	178 LGAPSGHLMRSTDAVMTSEVTJAPNPGNAQDPNDISGVASDNGITWTFDSITGG 237					
QY	248 LGQAKTIFVGVADPNPNPFWMSRSDGATWQAVGAPPTGFI PKAGVDPVNVLYATSN 307					
DB	228 AGTATRTLYGVADKEMNAVYRSTDGATWERLAGOPTGLAKGVDAANGVLYAYST 297					
QY	308 GGPYDSSGDVWKFVSTGVTWRISPVSTDTANDYFGYSGLTIDRQHNTIMVATQISW 367					
DB	298 GGPYDGGKRLRVATATGWTDISPAABADT---YGFSGGLTVDRQRGVMAATAYSSW 354					
QY	368 WPDITIFSTDSGATWTRIMWTSTYPNLSLRVYLDISAEPULTFGVQNPVPSPRLGMM 427					
DB	355 WPDITIFSTDSGATWQMSYTSYDRENRYTMVSSSPMLTWGAPNPAPPGTGLGMM 414					
QY	428 DEAMADPNDRMLKGCATLYATNDLTKW--DSGGQIHIAVMYKLESTAVNDLISPS 486					



Db 415 TEALEIDPDPDSRMVGTGATVYGTENLTNNMDEBGGTFAVEPMVRGLLEETAANDLASPPS 474  
Qy 487 GAPLISALGDLGGFTHADVTAVPSTIFTSPTFTTGTSDVYAEALNPSTIYVAGSPDPSSQP 546  
Db 475 GAPLISALGDLGGFTHADVTAVPSTIFTSPTFTTGTSDVYAEALNPSTIYVAGSPDPSSQP 546  
Qy 547 NDRAHAFSTDDGKNNWQSGEPGVTGTVAAASADSRFVWAPGDQPGQPVVAVVFGNSM 606  
Db 533 ---HIAFSTDNANMNGGTDPSGSGGTVAAGADSRFVWSP--EGAGVQYTTGFTSM 587  
Qy 607 AASQGPANAQIRSDRVNPKTFYALNGTFFRSTDDGVTPEQVAA--GLPSSGAVWFMHA 665  
Db 588 QASTLPALAIYSDVNRNATFYGFSGRFFVSTDDGATFTASAATLPAGD--GVAFKA 645  
Qy 666 VPKGEGDLMLAAS-----GLYHSTNGSSWSAITGVSSAVNVFGKAPGSSYPAVFV 720  
Db 646 LFGGSDVWLAGGAAGDPYGLMHSITDGGFTFRLPGVDADTVGFGKAAFGASQYTLFTS 705  
Qy 721 GTTGGVTAAYRSDDCCTTWTVLINDDOHYGNMGOAITGDHANLRVYIGTNGGIYVGD 780  
Db 706 AEIIGVVRGIFRSTDAATWTRVNDDAHQMWTGAATGDPVYGRVYVATNGRGIYGD- 764  
Qy 781 GGAPEGSPSPSVSPSASP 840  
Db 765 -----TSDTGGGTDPGPGDPPTP----- 782  
Qy 841 SASP 900  
Db 783 -----TGACEVTVYTNQWPGGQF--ADVRLTNTGT 811  
Qy 901 S-----SVVLS---TVYVRYWFTT--DGSSSTLVNCDM-----AATCGGIRAR--- 939  
Db 812 SAMNGWSTLWSPFGGQEVTRMNAEHTQAGTSVTAARVGNAGVAPGASVGFPGTSGRS 871  
Qy 940 -----SFGSVNPAATPTA 951  
Db 872 TAAEPGEFAVAGACPTA 889

## RESULT 2

D97013  
probably secreted sialidase, several Asp-boxes and dockerin domain [imported] - Clostrid  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 09-Jul-2004  
C:Accession: D97013  
R: Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
J.; Bacteriol. 183, 4823-4838, 2001  
J: Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cl  
A:Reference number: A96900; PMID:21359325; PMID:21359325  
A:Accession: D97013  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-839 <KUR>  
A:Cross-references: UNIPROT:Q97KKO; GB:AE001437; PIDDN:AAK78895.1; PID:gl5023820; GSPDB:C  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC0919

Query Match 39.7%; Score 2037; DB 2; Length 839;  
Best Local Similarity 49.5%; Pred. No. 1.2e-83;  
Matches 392; Conservative 118; Mismatches 252; Indels 30; Gaps 16;  
Qy 13 SRRVLSLAA-----TASPAVAAALGVLPIATSPAAAT-TOPYTWSNVAIGGGGVD 66  
Db 2 NKRIVSNVAGLSITFTTG-----VTHISAANKAAASVQSGYKMDNAKIGAGIYV 54  
Qy 67 GIVNFGAPGILYVTRDIGMYRMDANGRMIPLLDWVGNMNGYNGVSIADPINTK 126  
Db 55 AVIFNTEKOLIVARTMGAYRMDKANMKIPIFD--GFSDMTMLGCESIATPDIIDNR 112  
Qy 127 VAAVGMVITNSWPNDCAILRSSQAGTWTQITPPLKLGMMPRGMRERLAUDDNNNI 186  
Db 127 VAAVGMVITNSWPNDCAILRSSQAGTWTQITPPLKLGMMPRGMRERLAUDDNNNI 186

Db 113 VYIAGLYTNDMDENAYILSSODKNTWKRYQLPFKVGNMPPGRNGERLQIDENDKI 172  
Qy 187 LYFGAPSGKGLMRSTDSGATMSQMTNPPDVGTIYANPDTLGYOSDIGVWVAFDKSSS 246  
Db 173 LYFGAPSGKGLMRSTDSGATMSQMTNPPDVGTIYANPDTLGYOSDIGVWVAFDKSSS 246  
Qy 247 SLQASRTIFGVAD--PNNPVFMSRDGATQAVGAPGTGRIPHKGVDPVNHVLYATS 305  
Db 231 TKSGPQTQMYGAADKTKGNNTIYVINDGKMTSAKQKGLPHHGL-ABDGMLYISVS 289  
Qy 306 NTGPPYDSSGQVWKFVSTSGTWRISFVSTDTANDFYSGLTIDRQHPNTIYATQI 365  
Db 290 NTGPPYDSSGQVWKFVSTSGTWRISFVSTDTANDFYSGLTIDRQHPNTIYATQI 365  
Qy 366 SWWDPTIIFRSTDGATWTRIMDTSYNRLRYLQISAPRWLTFG--VQNPVPVPSPL 424  
Db 347 RWWDEEILYRSTDGATWTRIMDTSYNRLRYLQISAPRWLTFG--VQNPVPVPSPL 424  
Qy 425 GMDNEMADPENSDBMLYGTATLYATNDLTAKMDSGGQIHAIPWVKGLEETAANDLISP 484  
Db 407 GMDNEMADPENSDBMLYGTATLYATNDLTAKMDSGGQIHAIPWVKGLEETAANDLISP 484  
Qy 485 PSAPLISALGDLGGFTHADVTAVPSTIFTSPTFTTGTSDVYAEALNPSTIYVAGSPDPSS 544  
Db 467 TKGAQLISAVGDDCGFYHDDITKVPSSKMTTPNFSAITSIYAESVPMFVVRGVNDTSK 526  
Qy 545 QPNDRHAFSTDDGKNNWQSGEPGVTGTVAAASADSRFVWAPGDQPGQPVVAVVFG 603  
Db 527 NQDKDCGISYDGGKNNWSPAGSNTISGVYKAGTVAAGAAKTIWSP--EGANAAASTDNG 585  
Qy 604 NSWASQGPANAQIRSDRVNPKTFYALNGTFFRSTDDGVTPEQVAA--GLPSSGAVWFM 662  
Db 586 NKRIVSNVAGLSITFTTG-----VTHISAANKAAASVQSGYKMDNAKIGAGIYV 54  
Qy 663 FHAVGKEGDLMLA--ASSGLYHSTNGSSWSAITGVSSAVNVFGKAPGSSYPAVFV 720  
Db 644 FKTYIGHGDIWAGKXGLMHSITDGGFTFRLPGVDADTVGFGKAAFGASQYTLFTS 705  
Qy 721 TIGGVTAAYRSDDCCTTWTVLINDDOHYGNMGOAITGDHANLRVYIGTNGGIYVGD 780  
Db 704 TIGGVTAAYRSDDCCTTWTVLINDDOHYGNMGOAITGDHANLRVYIGTNGGIYVGD 780  
Qy 781 GAPSPPSPSVS 793  
Db 764 GS-QPTPSPSVT 774

## RESULT 3

T00349  
Avicelase III - Aspergillus aculeatus  
C:Species: Aspergillus aculeatus  
C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 09-Jul-2004  
C:Accession: T00349  
R: Arai, M.; Takada, G.; Kawaguchi, T.; Sumitani, J.  
submitted to the EMBL Data Library, June 1998  
A:Description: Avicelase III from Aspergillus aculeatus.  
A:Reference number: Z14141  
A:Accession: T00349  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-856 <ARA>  
A:Cross-references: UNIPROT:O74170; EMBL:AB015511; NID:d1199887; PID:d1029971  
C:Genetics:  
A:Gene: avIII  
F:823-854/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 33.9%; Score 1740; DB 2; Length 856;  
Best Local Similarity 43.5%; Pred. No. 2.1e-70;  
Matches 366; Conservative 139; Mismatches 297; Indels 40; Gaps 18;  
Qy 30 AALGVLPIATITASPPAAATOPYTWSNVAI--GGGFFVGGIVFENEGAPGILYVTRDIGMY 88  
Db 4 SLLALCALGLKADDAASQAYTKNVVTVGGGCGFTPTQIVNPSAKGVAYARTDIGAY 63



QY 89 RWDANGREWILLDMVG---WNNWYGVVSIADPINTNKVMAAVGMYTNSWDNDGAI 145  
DB 64 RLN-SDDTWTPLMDWVGNDDTHWDM---GIDALATDPVTDRIYVAVGYTINWDNVGSI 119  
QY 146 LRSSDOGATWQITPLPFKLGCMRPGRMGERLAVDPNNNDNLIYFGAPSGKGLWRSTDGSA 205  
DB 120 LRSTDQGTWTETKLPFKVGVGNMGRGERLAVDPNKNLSIYFGARSGHGLWKSTDYGA 179  
QY 206 TWSQMTNPPDVGTYIANPTDTGTGYSQSDIQGVVWVAFDKSSSLGQASKTIIFVGVADPNP 265  
DB 180 TWSNVTSTFTGTTFQDSSST--YTSDDPGIAWVTFDSTSGSSGATPRIIFVGVADAGKS 237  
QY 266 VFWSDGATWQAVPGAPT-GFIPHKGVDPVNVHLYIATNTGPGYDSSGDWVKFSVT 324  
DB 238 VFKSEDAGATWAVSGEPQYGLPHKGLVSPBEKTLXIYSANGAGPYDGTNGTVHKYNI 297  
QY 325 SCTWTRISVPSTDTANDYFVSGLTIDRQHPNTIMVATQISWPDITIIFRSTDGATWT 384  
DB 298 SGVWTDISP---TSLASTYTGUGLSVDLQVPGTLMAALNCWNPDELIFRSTDGATWS 354  
QY 385 RIWDTWTSYPNRSLRYLDISAEPLMTFGVQPNP-EVPSPKLGMDMAIDPFNSDRMLY 443  
DB 355 PLWEKNGYPSINYYSYDISNAPWQDITSTDPFV---RVGMVYEAALAIIDPFDSNHWLY 411  
QY 444 GTGATLYATNDLTWDSGGQIHIAPMVKGLBERTANDLISPPSGAPLISALGDLGFTHA 503  
DB 412 GTGLTVYGGHDLTNWDSKHNVTVKSLAVGIEEMAVLGLITPPGGPALLSAVGDDGGFYHS 471  
QY 504 DTVATVPSTIFTSPVFTTGTSDVYAEALNPSIIVRAGSFDPSSOPNDRHVAFTDGGKNWFQ 563  
DB 472 DLDAAPNOAYHTPTGTGTNGIDYAGNKPNSIIVRSGASO--DYPT---LALSNFGSTWYA 526  
QY 564 GSEPGVTTGGTVAASADSRVWAPDGPQVAVVAVFGNSWAAASQGVPAANAQIRSDRV 623  
DB 527 DYAASTSTGTGAVALSADGDTVLLMSSTSGALVSKSQG---TLTAVSLSPGAVIASDKS 583  
QY 624 NPKTFYALNSGTYFRSTDGVTQFVPAAGLPSGSAVGMFHAVPKGBGLMWLAASGLVH 683  
DB 584 DNTVYGGSGAIYVSKNTATSTFTKTVS-LGSSTTVNAI-RAHPSIAGDVMASTDKGLWH 641  
QY 684 STNGSSWSAI-TGVSSAVNVFGKSPGSSYPVAVVGTGIGVTVGAYRSDCGITWLI 742  
DB 642 STDYGSTFTIGSGVGTAGMSFGKASGTGSVWYIGFTIDCAAGLPKSEADAGTNQVI 701  
QY 743 NDDQHOYGNWQA--ITGDHANLRVYIGTNGRGIYVGDIGAPSG-----SPSPSV 792  
DB 702 SDASHFGSG-GSANVVGDLQTYGRVGRHERPGHLLRQSRPAGRHGDGDDTTYSKT 760  
QY 793 SPSPASPSLSPSPSSSPSPSPSPSSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 852  
DB 761 STTVSTLLKTTSSASTSSSTTVKTTSSSTTSKASSTTTTITTTTSSGTTATA 820  
QY 853 SS 854  
DB 821 SA 822

## RESULT 4

F72393  
hypothetical protein TM0305 - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 09-Jul-2004  
C:Accession: F72393  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;  
C.M.

Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A:Reference number: A72200; MUID:99287316; PMID:10360571  
A:Accession: F72393  
A:Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-707 &lt;ARN&gt;

A:Cross-references: UNIPROT:Q9WYEL; GB:AE001712; GB:AE000512; NID:G4980799; PIDN:AAD353

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TM0305

Query Match 22.6%; Score 1161; DB 2; Length 707;  
Best Local Similarity 35.2%; Pred. No. 9.9e-45;  
Matches 264; Conservative 123; Mismatches 275; Indels 89; Gaps 26;

QY 52 YTWSNVATGGGFGVGDIVNEGAPGLIYRTDIGMYRWDPAANGRWIPLLDWGVNHWY 111  
DB 21 FEWKSVEINGGGFVPGIIIFHPASGLLYARTDVGGLYRWDDEETKRMKQLDFLRDDSDY 80  
QY 112 NGVVSIAADPINTNKVMAAVGMYTNSWDNDGAILRSSDOGATWQITPLP---FKLGNM 168  
DB 81 MGVLVALDPSPKRIYAMTKYTQDW-AGYGAILSEDYGETWIVNLKYGIKVGNE 139  
QY 169 PGRGMGERLAVDPNNNDNLIYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDTT 228  
DB 140 DGRNAGERLOVDNPFSSVLFMGT-TKYGLWKSEDFGKNWKKVDSFPST----- 186  
QY 229 YQSDIQGVVWVAFDKSSSLGQASKTIIFGVADPNPVPFWSRGGATWQAVPGAPTGP 288  
DB 187 -----SVTFVLDFEKSGEKSPTRIFVGCSEPKG-IFVTEDGGITWNLPLNDLIP 239  
QY 289 HKGVDPVNVHLYIATNTGPGYDSSGDWVKFSVTGWTTRISPVSTDTANDYFVSG 348  
DB 240 LRGLKH--DGILYVTLNALGNPAGTGAVMKXVIADQKWDYVTPMKGD-----FGYCG 291  
QY 349 LTIQRHPNTIMVATQISWPDITIIFRSTDGATWTRIMDWTSYPNRSLRYLDISAEPW 408  
DB 292 IDVQE---NVIVSTLDRWYPHDEIFISLNGGETWRLLEKANF-----DINKAPW 339  
QY 409 LTFVQPNPVPSPKLGMDMAIDPFNSDRMLYGTGATLYATNDLTWDSGGQIHIAP 468  
DB 340 IK---DLNP-----HWISD-VKIDPDMNRAIFTTGYGVWVYELKKSPEG---MGK 384  
QY 469 MYK-----GLEETANDLISPPSGAPLISALGDLGFTHADVTAVPSTIFTSPVFTGT 522  
DB 385 PVKWIPENRGLEETVVLQVLPPIGERPLLSAIDWGGFRHESLDTPTSSMY-KPLKWTSL 443  
QY 523 SDVYAEALNPSIIVRAGSFDPSSOPNDRHVAFTDGGKNWFQ-GSEPGVTTGG--TVAA 579  
DB 444 GIATAYQNSKFVARVHTY---TYP---FLSYSDGGINWRRIETVPEGITDGGRLS 497  
QY 580 ADGSRFVWAPDGPQVAVVAVFGNSWAAASQGVPA---NAQIRSDRVNPKTFYAL--SNG 634  
DB 498 NDKTILWSPAN--HEVIVSSDKSKWKKAISVPVPFNPFPASDPVNPFSKYIFDWKNG 555  
QY 635 TFYRSTDGVTQFVPAAGLPSGSAVGMFHA-----VPGKGDLMWLAAS-SGLYHSTNG 689  
DB 556 DPLISKDGGKSPMK-GAKLPSPDNWVSLYSFPFLAPDREGDIWLALQWNLGYRSKDG 614  
QY 690 SWSAITGVSSAVNVVGRKSPGSSYPVAVVGTGIGVTVGAYRSDCGTTLINDDOHOY 749  
DB 615 TFERLGNVDIAYVIGFGAPKEGTDYPAIYLNGMVNVGYGIFWMTDEGKTMWRINNRK 674  
QY 750 GNMGQAITGDHANLRVYIGTNGRGIYVGD 780  
DB 675 G-WIHYMIGDMNEFGRIPLGTGREGIIVGEV 704

## RESULT 5

T31085

xylanase - Caldicellulosiruptor sp.

C:Species: Caldicellulosiruptor sp.

C:Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 09-Jul-2004

C:Accession: T31085

R:Morris, D.D.; Gibbs, M.D.; Ford, M.; Thomas, J.; Bergquist, P.L.

submitted to the EMBL Data Library, December 1997

A:Description: Family 10 and 11 xylanase genes from Caldicellulosiruptor sp. Rt69B.1.

A:Reference number: Z20972







Db	253	KWACHWDDVHNGAAL---LLAKITGKIYKQIIIESHLDVWITGNGERIKYTPKGLAWLD	309
Qy	564	---GSPGEGVTTGGTVAASAD-----GSRFVWAPGDPGPVVVAVGFG--	603
Db	310	QWGLRYATTTAFLAVFVSDWVGCPSKTEIYKFGESQIDYALGSAGRS--FVVGFGTN	367
Qy	604	-----NSWAASQGVPA-----NAQIRSD-----	621
Db	368	PPKPRPHRTAHSSWADSQIPSYHRTHTLYGALVGGPGSDSDYDDISNVNNEVACDYNA	427
Qy	622	-----RVNP-----KTFYALSNCTFYRSTDGVTGTFQPVAAAGLPSSGAVGWMPH	664
Db	428	GFVGLAKMYQLYGGNPIDFKAIETPTNDEFF-----VEAGINASGTNFIK	476
Qy	665	AVPGKEGDIWLAASS-----GLYHSTNGSSWSAITGVSSAVNVGFGKAP-----	710
Db	477	AIWNQSG-WPAKATDKLKFYFVDLSLSEIKAGYSPNQLT-LSTNTNQAKVSGPVVMDA	534
Qy	711	---GSSYPVAVVGTI---GGVTGAYRSDDCGTTWVLINDDOHQYGNWGOAITGDHANLREV	766
Db	535	SKNIYYILVDTGTLIYGGQDKYKKE-----VQFRIAAPQNVQNDNSDYSFQDIKGV	588
Qy	767	YIGT-----NGRGIVYGDIGGAPSGSPSPSVSPASPSLSPPSPSPSPSPSP	816
Db	589	SGSVVVKTKIPILYDGDVKGWGGPGTSGAIPPTA-----TATPTPTPTPTPTPTPT	642
Qy	817	SSPSSSP	875
Db	643	TS--TATPTPTPTPTPTPTPTPTA--TPTSTPTTS-----TPSSTPVAGQIKV	690
Qy	876	QYKNDNSAPGDNQIKGLQVNTGSSVDLSVTYVYVTRDGGSTLVNCDWAAIGCG	935
Db	691	LYANKETNTTIRPWLKVNTGSSIDLRSVIRYVYVTRDGGSTLVNCDWAAIGAS	749
Qy	936	NIRASPGSVNPATPTADTYLQ 956	
Db	750	NVTFKFEVLSSVSGADYYLE 770	
RESULT 7			
Tl7120			
cellulase (EC 3.2.1.-) precursor, thermoactive - Caldocellum saccharolyticum			
C;Species: Caldocellum saccharolyticum			
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004			
C;Accession: Tl7120; A43745			
R;Te'o V.S.; Saul, D.J.; Bergquist, P.L.			
Appl. Microbiol. Biotechnol. 43, 291-296, 1995			
A;Title: Cella, another gene coding for a multidomain cellulase from the extreme thermoph			
A;Reference number: Z18698; MUID:95336703; PMID:7612247			
A;Accession: Tl7120			
A;Status: preliminary; translated from GB/EMBL/DBJ			
A;Molecule type: DNA			
A;Residues: 1-1742 <TEO>			
A;Cross-references: UNIPROT:P22534; EMBL:L32742; MID:9537499; PID:9537500; PIDN:AAA91086			
R;Luethi, E.; Bhana Jaemat, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.			
Appl. Environ. Microbiol. 57, 694-700, 1991			
A;Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene coding			
A;Reference number: A43745; MUID:91247819; PMID:2039230			
A;Accession: A43745			
A;Status: preliminary			
A;Molecule type: DNA			
A;Residues: 1516-1544, 'A', 1546-1742 <LUE>			
A;Cross-references: EMBL:M36063; MID:g144292; PIDN:AAA72860.1; PID:g144293			
C;Genetics:			
A;Gene: Cella			
C;Keywords: glycosidase; hydrolase			
Query Match 7.7%; Score 397.5; DB 2; Length 1742;			
Best Local Similarity 21.7%; Pred. No. 2.7e-10;			
Matches 189; Conservative 109; Mismatches 269; Indels 303; Gaps 35;			
Qy	262	PNNVPWRSRGGATWQA----VEGAPTGFIPHKGVDFPVNHLVY-IATNSTGPGYDGGSG	316

Db	48	PN-----WVRN-----NWRGDSALKDQDNGLDLTCGFWFADGHDVKFNLPMSTYGTMLSWAA	100
Qy	317	DWVKFSVTSGTWRISPPVSTDTANDYFCYSGLTIDROHPNTIMVATQI-----SWM--	368
Db	101	EYKDAFVKSQGLEHI--LNQIEWNDYF-----VKCHPSKVVYVQVGDGKDHAWG	151
Qy	369	-----PDTIIFRSTDGGATWTRIMDWTSPNRSRLRYLVDISAEPMLTFCGVQNPVP	419
Db	152	PAEVQMREPSKVTQSSPGSA-----VVAETAASLAASIVLKDRN	193
Qy	420	PSPKLGWMDMAIDPFNSDRMLYGTGATLYATNDLTWKDSCGQIHIAPMVKGLEETAVN	479
Db	194	PTKAATYLGHA-----KOLYBFAEVTKSDSG-----YTAAN	224
Qy	480	DLISPPSG-----APLISALGDLGGFTHADVTAVPSTIFTSVFTTGTSTVDY---	526
Db	225	GYNWSGDFYBELSMAWVLYLATNDSTYLTAKB-----SYQNWPKISGNSNIIDYKWAH	279
Qy	527	-----AELNPSIIVRAGSPDPSPQPNDRHVAFTDG-----GKNW-----	561
Db	280	CWDDVHNGAAL---LLAKITDKYQIIIESHLDYWTGNGERIKYTPKGLAWLDOWGS	336
Qy	562	-----FQSGEPGGVTTGGTVAASADG--SREVNAPGDPGPVVVAVGFG-----	603
Db	337	LRYATTTAFLAVFVSDWSCPTGKKEKYKFGESQIDYALGSTGRS--FVVGFGTNPCKR	394
Qy	604	-----NSWAASQGVPA-----NAQIRSD-----	621
Db	395	PHHRTAHSSWADSQIPSYHRTHTLYGALVGGPGSDSDYDDISNVNNEVACDYNA	454
Qy	622	-----RVNP-----KTFYALSNCTFYRSTDGVTGTFQPVAAAGLPSSGAVGWMPH	668
Db	455	ALAKMYLLYGGNPIDFKAIETPTNDEFF-----VEAGINASGTNFIKAI--	501
Qy	669	KEGDLWLAASGLYHSTNGSSWSA-----ITGVSSAVNVGFGKSGAPSSYPAVF	718
Db	502	-----VNNQSGWPARATNKLKFRYFVLDSELIKAGYS-----ENQL	537
Qy	719	VVGTI---GGVTGAYRSDDCGTTWVLIND-----DHOYGNWGOAITGDHANLRR	765
Db	538	TLSTNTNQAKVSGPVVWDSSRNIIYILVDTGTLIYGGQDKYKKEVQFRIAAPQNVQ--	596
Qy	766	VVGTNNGRIVYGDIGGAPSGS-----PSPSVSPASPSLSPPSPSPSPS	807
Db	597	---WDSNDYSFQDIKGVSSGVVTKYIPLYDEDIKVMGEEFGTS-----GVSPTPTAS	648
Qy	808	SSP	867
Db	649	VTPPTPTPTA-----TPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT	700
Qy	868	P-VSGGVKVOYKNDNSAPGDNQIKGLQVNTGSSVDLSVTYVYVTRDGGSTLVN	926
Db	701	PATSGQIKVLYANKETNTTIRPWLKVNTGSSIDLRSVIRYVYVTRDGGSTLVN	760
Qy	927	CWMAALGCGNIRASFGSVNPATPTADTYLQ 956	
Db	761	-DMAQIGASNVTPFKVKLSSSVSGADYYLE 789	
RESULT 8			
SS0755			
hypotheical protein Vsp-3 - Chlamydomonas reinhardtii			
C;Species: Chlamydomonas reinhardtii			
C;Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 09-Jul-2004			
C;Accession: SS0755			
R;Woessner, J.P.; Molendijk, A.J.; van Egmond, P.; Klis, F.M.; Goodenough, U.W.; Haring			
Plant Mol. Biol. 26, 947-960, 1994			
A;Title: Domain conservation in several volvoclean cell wall proteins.			
A;Reference number: SS0754; MUID:95093034; PMID:8000007			
A;Accession: SS0755			
A;Status: preliminary			
A;Molecule type: mRNA			



A;Residues: 1-473 <WOE>  
A;Cross-references: UNIPROT:Q99620; EMBL:L29029; NID:g530875; PIDN:AAB53953.1; PID:g53087

Query Match  
Best Local Similarity 33.0%; Score 372.5; DB 2; Length 473;  
Matches 132; Conservative 36; Mismatches 139; Indels 93; Gaps 17;

Qy 492 SALGDLGGFTHADVAVPSTIFTSPFTTGTSDVDYAEIENPSIIIVRAGSDFDPSQP-----547  
Db :|||:|||:|||||||:|||||||:|||||||:|||||||:  
30 SAAGPYTVVITYANSQVIASLTRLSLV---ETNLKY--LTPGGLARKNPITQTTSQPSAAQL 84  
Qy 548 ---DRHVAESTDGGKNW---FCGESEPGGVTTGCTVAASADGSRFWA PGDPGPVVYAVG 601  
Db :|||:|||:|||||||:|||||||:|||||||:|||||||:  
85 AGADAAYILLSNRYYSWDTEKMGSSPNKV-----RDLDXNVNAGGS----LVLDDG 131  
Qy 602 FGNSWAASOCVPANAOIRSDRVNPKTFYALLSNG--TFYIRSTDCGGVTFPVAA GLPSSGA V 659  
Db :|||:|||:|||||||:|||||||:|||||||:|||||||:  
132 YSTA GTNTFEVQLIDAVLTGKAGSGCTGALYGNVNVRANS SFGKITSLPVKGS- 190  
Qy 660 GVNFHAVPGKEGDMLAAASGLYHSHTNGSSSAITGVGSVANVGFKSA---PGSSSYPA 716  
Db :|||:|||:|||||||:|||||||:|||||||:|||||||:  
191 -----SRGESLGTCTGSAVLFSNPDKWTKA VT--ASAITSVSGKGAITFIGSSFAM 241  
Qy 717 VFVVFVTIGVTGYNRSDDCGTTWLINDDOHQYNWGCOALTGDHANLRRIYIGTG NG RIV 776  
Db :|||:|||:|||||||:|||||||:|||||||:|||||||:  
242 PHLKG-----YEDMS-----GVAVT--LANFNKT-----263  
Qy 777 YGDIGGAPS-----GSPSPSVSPA SLSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 830  
Db :|||:|||:|||||||:|||||||:|||||||:|||||||:  
264 ----GASPSPSPKASPSPKVSPSPSPKASPSPFKASPSPSPFKASPSPSPFKASPS 319  
Qy 831 SPSPSPSPSPASPSPPSP 868  
Db :|||:|||:|||||||:|||||||:|||||||:|||||||:  
320 SPSPSPSPSPKASPSPSPSVQPSASKPSPSPSPSPS 359

RESULT 9  
A49227  
sialidase - Actinomyces viscosus  
C;Species: Actinomyces viscosus  
C;Date: 19-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004  
C;Accession: A49227  
R;Yeung, M.K.  
Infect. Immun. 61, 109-116, 1993  
A;Title: Complete nucleotide sequence of the Actinomyces viscosus T14V sialidase gene.  
A;Reference number: A49227; UID:93114861; PMID:8418033  
A;Contents: T14V  
A;Accession: A49227  
A>Status: preliminary  
A:Molecule type: nucleic acid  
A;Residues: 1-901 <YEUC>  
A;Cross-references: UNIPROT:Q44562; GB:L06898; NID:q289087; PIDN:AAA21932.1; PID:g141852  
A;Note: sequence extracted from NCBI backbone (NCBIN:121598, NCBIPI:121599)

Query Match  
Best Local Similarity 22.0%; Score 346.5; DB 2; Length 901;  
Matches 228; Conservative 99; Mismatches 304; Indels 405; Gaps 51;

Qy 30 AALGVLPPIAIT-----ASPHAHAATTPQYTWSNVAIGGGGFVDGIVNEGAPGI-LYVRTD 83  
Db 14 ALLGSLPLAATGATGLIAAAPAHAVPT-----SDGLADVTTIQVNAPADGLYSVG D 62  
Qy 84 I-----CGMYRWDAANGRWIFLLDWGNWNWYGNGVVSIAADRPINTKWAAV G 132  
Db :|||:|||:|||||||:|||||||:|||||||:|||||||:  
63 VMTFNITLNTSGEASHYAPASTNLGNVSKCRWRN-----VPA GTTKT-DCTG 110  
Qy 133 MYTNISWDNDGAILRSSDQAATQQITPLPKLGMEGRGMGERLA VD PNNDNILYFGAP 192  
Db 111 L-----ATHVTVAEDLKAGGFTQIAY-EVKAVE-----Y 139  
Qy 193 SGKGL-WRSTD SGATWSQMNTFFDVGYTI ANPT-----DTTYGDSDIQG VVWAFDKS 244  
Db 140 AGKALSTPETIKCATSPVKANSRLRSVISITPSSSQNYKLGDVTYTVRVS V-----191



A; Cross-references: UNIPROT:P22533; GB:L01257; NID:g144290; PIDN:AAA71887.1; PID:g144291  
A; Note: sequence extracted from NCBI backbone (NCBIN:121576, NCBI:P121577)  
R; Luethi, E.; Bhana Jasmal, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.  
Appl. Environ. Microbiol. 57, 694-700, 1991  
A; Title: Cloning, sequence analysis, and expression in *Escherichia coli* of a gene coding  
A; Reference number: A43745; MUID:91247819; PMID:2039230

A/Accession: U01858  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-337; 'ppRQHQRHQ' <LUE>  
A/Cross-references: EMBL:W36063; NID: g144292; PIN: AAT72861.1; PID: g144294  
A/Note: The authors translated the codon CAC for residue 262 as Glu  
A/Note: This sequence has been revised in reference A48954  
C/Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match	6.6%;	Score 340.5;	DB 2;	Length 1331;
Best Local Similarity	22.7%;	Pred. No. 6.8e-08;		
Matches 140: Conservative	97;	Mismatches 189;	Indels 191;	Gaps 29;

QY 426 WMD E A M A I D P F N S D R M L Y G T G A T I Y A T N D L T K W D S G G Q I H I A P M V K G L E E T A V N D L I S P P 485

486 SGAPLISALGDLGGFTHADVTAVPS--TIFTSPVFTTGTSDVYAEINPSIIVRA---GSF 540

Db 179 ---AIKALRD-AGFKHTIMVDAFWGQDWSNTMRDQAQIEMADPLRLNVFSIHMGVY 233

DB            : : : : : : : : : : : :  
234 NTASKVEEYIKSFVDKGLPLVIGFHHQHTD-----GDPDEE 270

QY	596 W-----YAVG-FGNSWAASQGVAPANAQIR\$--DRVNPKTIFYALSNGTTFYKSTIDGG---64
P4	271 AITMAVAKVQVCIIGWISWGANS\$VUCVIIMYNMNNDDNDDPW-----GWKYNTAIGTSS325

QY 644 -----VTQPVAAGLP-----SSGAVGVNFHAVPCKEGDL----- 673

DD	IPTPTSIVTIFPIFPIFIPIVIAIPIFIPIFVSIPAISSQINVLII	-ANNRNSNNINI
326		
CV	---WL-----AASSGL-----YHSTNGGSSNSAITGVIS--SAVNVE-----GKSAP 710	
674		

Db 383 IRPWLKVVNSGSSSIDLSRVITIRYWTVDGERAQSAISDWAQIGASNVTFKFKVLSVVS 442

Db 443 GADY--YLETGFKSAGLOPGKDTGEIQMRPNKDWSNVTNQNDWSWIQSWTS-----494

QY 765 RVYIGTNGRGIYGD----IGGAPGCGSPSPSVSPSASPSISLSPSPSSSSPSPPSPSSSP 820

[illegible][illegible]

Db 577 KETNSTNTIRPWLKVVNSGSSSIDLSRVTRIYWTVDGERAQSALS-DWAQIGASNTVF 635

QY 940 SFGSVNPAFTADTYLQ 956  
| : : | | :  
nb 636 KEVKTSSSSVSGADYVLE 652

## RESULT 11

hypothetical protein slr0442 - Synechocystis sp. (strain PCC 6803)  
C;Species: Synechocystis sp.

C;Accession: S76211  
C;Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 09-Jul-2004

o. K.: Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda

DNA Res. 3, 109-136, 1996  
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp.  
A;Reference number: S74322; MUID:97061201; PMID:8905231  
A;Accession: S76211  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-611 <KAN>  
A;Cross-references: UNIPROT:P74375; EMBL:D90914; GE:AB001339; NID:g1653477; PIDN:BAAL847  
Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 6.0%; Score 308.5; DB 2; Length 611;  
Best Local Similarity 26.7%; Pred. No. 7.3e-07;  
Matches 142; Conservative 51; Mismatches 180; Indels 159; Gaps 21;

QY 429 EAWA--IDPFNSDR-----MLYGTGATLYATNDLTKWDSCGGQTHIAPMWKGLESTAVNDLI 482

QY 483 S-----PPSGAPLISALG-----DLGGFTHADVTAVPSTIFTSVPFTTGTSDY 526

Db 207 TRLMVDIPISGKPGCTSTFGGSDTPALWISGGVDKGTAGDITDVTGTPAVPKTG----- 261

Db 262 -EIRGNIVLAGCGSTDLTDAYIDDLNSTQNAGGFAAVRASTPMPSPLSALASARSONR 320

QY	585	FWAPGDPQPVVYAVGFGNS	-----WAAASQVPAANAQIRSDKVNPAI	62
ST	201	WYMTNVCNNINST	STDPSCGEORVDGVVVVSSSI	368
		DRKPEINEDNTGGKK		

QY 628 FYALSNGTFYRSTGGVT-----FQPV---AAGLPSSCAVGVMFHAVPGKEGDLWLAAS 678

DB 369 YRLVITGILGRNAEINGICNIADGCGFIDIQIFGLSSSSGSI  
Q4 679 SGIYHSTNGGSSWSAITGVSSAVNVGFKSA PGSSYPAFVVTGIVGTGAYRSDDCGTT 738

Db 410 -----CMNGNSRRTKAFV-LAPNYDIGRTGNGS-----YEGMLLGKT--YAKGNCAV- 454

Dbb-----NNGKLAVTQT-----SQTWGSVPPDL--RPLTA<sup>P</sup>SPTITSSFW 491  
70 WYELNDQKGLDMPGALFCEHAKLFGSGEAGLVGGPSTPGIATSLA

[illegible]

**Qy**

**Qz**

837 SPSPASPPSSPSSPSSPSSPTSSSPys-----GGVKVQYKN 880

Db 552 SPSPSPSPSPSPSPSPSPSPPTVTNVQNKKACDDLGCTYSQSGNN 603

RESULT 12  
S75138  
Chemohomologous en (strain DCC 6803)

C,Species: *Synechocystis* sp.  
A,Variety: PCC 6803  
C,Date: 25 Nov 1997 #sequence revision 25-Nov-1997 #text change 09-Jul-2004

C;Accession: S75138  
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyazaki, Y.; Wada, T.; Watanabe, A.; Yamada, M.; Shimizu, S.; Shimizu, S.; Okumura, K.; Okumura, K.

A1: Sequence analysis of the genome of the unicellular cyanobacterium *Synechococcus* sp. 3, 109-136, 1996

A;Accession: S75138  
A;Reference: S75242; NC015700220-1  
A;Status: nucleic acid sequence not shown; translation not shown

A;Residues: 1-1749 <KAN>  
A;Cross-references: UNIPROT:P73032; EMBL:D90903; GB:AB001339; NID:gf652127; PIR:G00000

C; Genetics:  
A; Start codon: GTG







T34434  
hypothetical protein K06A9.1a - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T34434  
R:Geisler, C.; Gattung, S.  
submitted to the EMBL Data Library, December 1996  
A:Description: The sequence of C. elegans coamid K06A9.  
A:Reference number: 221525  
A:Accession: T34434  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2232 <GEI>  
A:Cross-references: UNIPROT:Q81FK6; EMBL:U0846; PIDN:AAC70890.1; GSPDB:GN00028; CESP:K06A9  
A:Experimental source: strain Bristol N2; clone K06A9  
C:Genetics:  
A:Gene: CESP:K06A9.1a  
A:Map position: X  
A:Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 2075/1

Query Match 5.7%; Score 290.5; DB 2; Length 2232;  
Best Local Similarity 21.8%; Pred. No. 2.1e-05;  
Matches 239; Conservative 114; Mismatches 428; Indels 315; Gaps 45;

QY 18 SLLAATASFAVAALGLVPIA---ITASPAHAATTPQYTMGNVAI--GGGPFVDGIVENE 72  
DB 235 SLGTSSPLPSISITSAIPAIASSASSPSAASSTPVVLSSITQSSSGTFPPSSVASSP 294

QY 73 GAPGILYVRTDIGMYRMDAANGRIPLLDVWGNWNGYGVVSTAAAD-----PINTNK 126  
DB 295 STVG-----STSG-----AASSSYATVSTIAGSTGSTITVPGSS 330

QY 127 VMAAVGYMTNSWDPNDAILR--SSDQAGTQITP-LPFLKLGNNMP---GRGMRERLAVD 180  
DB 331 --STIGSTPSASSSSSGTMTISGSGTGTVTVPVPGSSSTPASSTPIASSSPGSGTIVA 388

QY 181 PNNNLIYFGAPSGKGL---WRSTDSGATWSQMTNPFDPGVYIANPTDTTGYQSDIQGV 237  
DB 389 PGSSSTYGSSTPSASSSSSGTMTNSGSGTGTVTVPVSSSTFGSGSTPIASSSSSGSTVT 448

QY 238 WVAFDK-----SSSLGQAS-----KTFVGVADPNNPVFWRDGGATWQA 278  
DB 449 VSGSSSTYGSSTPSASSSSAGTASTISGSGSTATVPGSS-----SSVGSSTQSA 500

QY 279 VPGAPTGFIPHKGVDPVNVHLYIATNTGGPYDGGSDGVKXFSVTSGTWTIRISVPVSTD 338  
DB 501 SPSSPG-----TWSTVSGP--TGSTVTVPGSSTSPA--PSSSPNPSSS 540

QY 339 TANDYFGYGLTIDRQHPNTIMVATQISWHPDTIIFRSTDGATWT---RIWDWTSYNR 395  
DB 541 PAS-----TGSTITISGSSIIIVST-----VSGSTVSGTGTSTQSTLASSTATPGS 586

QY 396 SLRYVLDISAE-----PWLTFGVQPNP---PVPSKLGWMDAAMADPPNS 438  
DB 587 SSTVPSSSSPQSSQSPAPNTGTTTSPQTSQSPSPSMNPSSSTTGTSSQSTTITPEGSTA 646

QY 439 DRMLYGTGATLYATNDLKWDSGGQIHIAPMVKGLEETA VNDLISPPSGAPLISALGDL- 497  
DB 647 SSPTGSGTSTSVATEVTSQT-----VPSGSSLGQTSTNSSPSPSSLSPTSGMSTLT 700

QY 498 -----GQFTHADVAVPS-----TIFTSVPVTTGTSVDYAEINPSTIIVRAGS 539  
DB 701 SEPPSPSTQSSGAQSTLTTPSPNPQSTSSLESSTSGATTSSGSAGTTMTSPSQSSSVGS 760

QY 540 FDPSSQPNDRHVAFSTDGKWFQGS--EPG--GVTTGGTVAASADGSRFVWAPG----- 590  
DB 761 SQGSTSP-----AASTTSGEMTSQGSTQTPGSSVSTSAAILTSTQOSVSTNSPGSTVTRP 815

QY 591 -----DPGQPVVY-----AVGFGNSWAASQGVPAQAQIRSDRVNPKTFYALSNGTFFYRS 639  
DB 816 STVSGSTSGSTVTVGSTEASTSGSVASSSPAPSTSQ-----NPNPSTSGSSMITQS 869

QY 640 TDGGVTFQFVAAGLPSGAGVGMFHAVPGKEGDLMLAASSGLYHSTNGGSSWSA-----IT 695

Db 870 PYPQSTSPVESSTTPS-----PGSPGTTLTSTSPSPSQSTTIGSTQGSTSPGIS 919  
QY 696 GVSSAVNVGFGKSAFGSSYPVAVVGTIGGVTGAYRSDDCGTTWLVINDDQHOYGNWQQA 755  
Db 920 TTSEEMTSQGSTQTPGSGSTGTQSTVSDST----- 951  
QY 756 ITGDHANLRVYVGTNGRIGVYGDIGGAPSGSPSPSPSASPSPSPSPSPSPSPSPSPSP 810  
Db 952 -----SSGSTVTVGSTEK--SSSPISTQNTNPSTSSGSSMSTQTPQSSQ 995  
QY 811 SPSP--SPSSSPSSSPSP-----SPSPSPSP-----SRSP-----SPSASPSPSS 848  
Db 996 STSPVESSTSGATSSSGSPGTTLTSTSPSPSPSTIGSSQGSTSPVVTISQGSTETPGS 1055  
QY 849 SPSPSSSP-----SSSP 881  
Db 1056 TQSTVTKESTVSGSASSGSTATWGTSTGSGSTSPNPQSTSPSTSGA-----T 1108  
QY 882 SAPGDNQIKPGLQVNVNTGSSSVDLSTVTYVYWFTRDGGSSSTLVYNC--DWAALCGGNIRA 939  
Db 1109 SPFGSS-----GTTLTISPSQSGSTIG-----SSQGSTSPVVTSGDMTSGSTQIPG 1159  
QY 940 SFGS--VNPATPTADT 953  
Db 1160 STGSTVTPQSTGSGST 1175

Search completed: October 5, 2005, 07:57:01  
Job time : 32.555 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 07:44:57 ; Search time 118.987 Seconds  
(without alignments)  
4118.611 Million cell updates/sec

Title: US-09-917-376-1  
Perfect score: 5135  
Sequence: 1 MDRSENELTWRSRLVSL.....RASFGSVNPATPTADTYLQX 957

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 03.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2479.5	48.3	882	Q82M04	streptomyc
2	2436	47.4	890	Q86727	streptomyc
3	2420.5	47.1	996	Q9AQH0	caldicellul
4	2037	39.7	839	Q97KK0	clostridium
5	2007	39.1	842	Q70DK5	clostridium
6	1931.5	35.7	857	Q8TFP1	aspergillus
7	1784	34.7	806	Q9P4T8	agaricus bi
8	1740	33.9	856	Q74170	aspergillus
9	1625.5	31.7	838	Q729M8	trichoderma
10	1587.5	30.9	861	Q7S1W6	neurospora
11	1442	28.1	739	Q82K30	streptomyc
12	1173.5	22.9	751	Q8FLM5	xanthomonas
13	1161	22.6	707	Q9WYE1	thermocoga
14	1151.5	22.4	751	Q8P9U5	xanthomonas
15	1125	21.9	812	Q8J0D2	geotrichum
16	993	19.3	776	Q764N8	geotrichum
17	431.5	8.4	1779	Q52374	caldicellul
18	409	8.0	196	Q8J1H7	agaricus bi
19	409	8.0	1770	Q9X3P5	caldicellul
20	402.5	7.8	921	Q918L8	caldicellul
21	398.5	7.8	1711	Q96311	anaerocellu
22	397.5	7.7	1742	Q96311	anaerocellu
23	387	7.5	2014	Q7U3X4	caldocellum
24	385	7.5	1664	Q7X2U2	uncultured
25	372.5	7.3	473	Q39620	chlamydomon
26	360	7.0	577	Q6FSJ1	candida gla
27	354.5	6.9	790	Q6FX25	candida gla
28	350.5	6.8	997	Q924I1	bacillus sp
29	346.5	6.7	901	Q44562	actinomyces
30	340.5	6.6	1331	MANB	caldocellum
31	336.5	6.6	741	Q82QF2	streptomyc

32	335	6.5	911	2	Q72WN1	desulfovibr
33	334	6.5	1751	2	Q9AQG4	caldicellul
34	325.5	6.3	930	2	Q9RFX5	caldicellul
35	325	6.3	1000	2	O24820	thermophili
36	320.5	6.2	991	2	Q94C44	chlamydomon
37	318.5	6.2	875	2	Q9F2B0	thiobacillu
38	316	6.2	1091	2	Q8KKF7	paenibacill
39	308.5	6.0	611	2	P74375	synechocyst
40	301.5	5.9	1749	2	P73032	synechocyst
41	298.5	5.8	1154	2	Q7U5X7	synechococc
42	297	5.8	714	2	Q7U5X6	synechococc
43	297	5.8	1915	2	Q9RPL0	acetivibrio
44	294	5.7	474	2	Q86154	dictyosteli
45	293	5.7	279	1	Y091_NPVP	orgyia pseu

## ALIGNMENTS

### RESULT 1

ID	Q82M04	PRELIMINARY;	PRT;	882 AA.
AC	Q82M04;			
DT	01-JUN-2003 (Tremblrel. 24, Created)			
DT	01-JUN-2003 (Tremblrel. 24, Last sequence update)			
DT	01-MAR-2004 (Tremblrel. 26, Last annotation update)			
DE	Putative endo-1,4-beta-glucanase.			
GN	Name=celA3; OrderedLocustNames=SAV1856;			
OS	Streptomyces avermitilis.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Streptomycineae; Streptomycetaceae; Streptomyces.			
OX	NCBI_TaxID=33903;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=NA-4680;			
RX	MEDLINE=22608306; PubMed=12692562;			
RA	Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,			
RA	Sakaki Y., Hattori M., Omura S.;			
RT	"Complete genome sequence and comparative analysis of the industrial			
RT	microorganism Streptomyces avermitilis.";			
RL	Nat. Biotechnol. 21:526-531(2003).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=NA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;			
RX	MEDLINE=21477403; PubMed=11572948; DOI=10.1073/pnas.211433198;			
RA	Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,			
RA	Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,			
RA	Kikuchi H., Shiba T., Sakaki Y., Hattori M.;			
RT	"Genome sequence of an industrial microorganism Streptomyces			
RT	avermitilis: deducing the ability of producing secondary			
RT	metabolites.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).			
DR	EMBL; AP005028; BAC69567.1; -.			
DR	HSSP; P07986; 1EXG.			
DR	GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.			
DR	GO; GO:0005375; P:carbohydrate metabolism; IEA.			
DR	InterPro; IPR001919; Bac celose-bind.			
DR	InterPro; IPR008965; Cellul bind.			
DR	InterPro; IPR002860; Glyco hydro_BNR.			
DR	InterPro; IPR011040; Sialidase.			
DR	Pfam; PF02012; BNR; 10.			
DR	Pfam; PF00553; CBM_2; 1.			
DR	SMART; SM00637; CBD_II; 1.			
KW	Complete proteome.			
SQ	SEQUENCE 882 AA; 92152 MW; A3C8E753CE69F13D CRC64;			

Query Match	48.3%;	Score 2479.5;	DB 2;	Length 882;
Best Local Similarity	50.1%;	Pred. No. 1.5e-96;		
Matches 476;	Conservative 129;	Mismatches 230;	Indels 115;	Gaps 15;
QY	11	MSRRRLVSLAATAATSPAAAL--GVLPITATSPAAHAATTQPYTSNVAIGGGGVVDGI	68	
DB	1	MRRTFR-----TAVLALAGLPAGTTPALAAAPTATIAADTYSWKNARVDGGGVFGI	55	



QY	69	VNEGAPGILYVRTDIGMYRWDAANGRWIPLLDWGNNMGNGYGVWIAADPINTNKVW	128
DB	56	VFNRSKLNLAARTDIGAYRWAESSKWTPLPLDSVGSWDGHTGCVSLASDSVDPNKY	115
QY	129	AAVGMNTNSWDPNGAGAILRSSDQAGATWQITPLPFKLGGMGPGRGMERLAVDPNNNDILY	188
DB	116	AAVGYTNTNSWDPNGAVLRSRDRGASQKTDLPFLKLGGMGPGRGMERLAVDPNNRNSVLY	175
QY	189	FGAPSGKLGWSTDSGATWSQWNPDPVGTIYANPTDITGVQSDIQGVVWVAFDKSSSL	248
DB	176	LGAPSGKLGWSTDSGASWSQVTPPNVGTYYQDATDSGVASDNQGIWVTFDESTGSP	235
QY	249	GOASKTIFVGVADPNPNVFWSRDGAATWQAVPGAPTGFPHKGVDPDPVNVHLYIATSNITG	308
DB	236	GSSRTFVYGVADKONSYYRSTDAATWSRLAGQPTGHLAHKGVLDAAAGCLYLAYSCKG	295
QY	309	GPYDSSGDVWKFVSTGTRISPVSTDTANDYFGYSGLTIDRQHPNTIMVATQISNW	368
DB	296	GPYDGGKQLWRYTKTGTTWISPVAEADT---YGFSGLTVDQRHPTGMATAYSSW	352
QY	369	PDTIIFRSTDGATWTRTMDWTSYPNRSIRYVLDISAEPLWTFGVQPNPPVPSPKLGWMD	428
DB	353	PDTQLFRSTDSGGTWTAKWDYTSYPSRNRFTMDVSSSGFWLTWGANPAPPEQTPLKGMWT	412
QY	429	EMALDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHIAPMVKGLEETAVNDLISPPS-G	487
DB	413	ESLEIDPFDSARMYGTGATYGTUNLWNDSGSQFTIKPMARGLEETAVNDLASPPSGG	472
QY	488	APLISALGDLGGFTHADYVAVPSTIFTSPVFTTGTSDVYAEINLPSIIVRAGSFDPSQPN	547
DB	473	AQLPSALGDIIGFRHTLDTTVPSLMYTSPNFTTSLDYAETDPTGVVRVGNLD--SGP-	529
QY	548	DRHVAFTDGGKNWFGQSPGCVTGGTVAASADSRFWAPCDPGQVPIVYAGFNSWA	607
DB	530	--HVAFTDNGANWFPAGADPSGVSGGTVAASADSRFWSPAGTG--VQYTTGFGTWS	585
QY	608	ASQGVPAANAQSRDRVNPKTFFVALNGTFFRSTDCGVTGVPVAA--GLPSSGAVGVWFMHAV	666
DB	586	ASAGAPAGAVESDRVDPKTFYFGSGRFPYVSSDGGATFTASATGLPSGDS--VRFRAL	643
QY	667	PKEGDLMALASS-----GLXSTNGSGSSAITGVSSAVNVGFGKSAPGSSYPAVFVVG	721
DB	644	PGTKGDIWLAGSGADGAGLWHSTDGGAAFTKLTATVDQADTIGFCKAATGASYQTLTSA	703
QY	722	TIGGVTGAYRSDCCTTWVLINDDOHQVGNWQATIGDHANLRVYIGTNGRGIVYGDIG	781
DB	704	KITGGVGRGIFRSTDKGASWTRVNDDAHQWGTGAATIGDPRVYGRVYVSTNGRGIVYGDTA	763
QY	782	GAPSGSPSPSVSPASPSLSPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPS	841
DB	764	GSSDGG-----	769
QY	842	ASPPSP	901
DB	770	-----GTPEAPT-----GACTVYIRITNQWSSGGFQ--ADVQLANTGST	805
QY	902	SVDLSTVTVRYMFTDRGSSSTLYVNCDMAIICG-----NIRASFGS	943
DB	806	AMDGWSLG---WSPGDBGVETQLWNASVAQAGSGVTAANLAWNGRVAAGS	852
RESULT 2			
O86727			
ID	O86727	PRELIMINARY; PRT; 890 AA.	
AC	O86727		
DT	01-NOV-1998 (TreeBLrel. 08, Created)		
DT	01-NOV-1998 (TreeBLrel. 08, Last sequence update)		
DT	01-MAR-2004 (TreeBLrel. 26, Last annotation update)		
DE	Putative secreted cellulase.		
GN	ORFNames=SC5C7.30c;		
OS	Streptomyces coelicolor.		
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;		

OC	Streptomycineae; Streptomycetaceae; Streptomyces.		
OX	NCBI_TaxID=1902;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=A3(2) / M145;		
RX	MEDLINE=21996410; PubMed=12000953; DOI=10.1038/417141a1;		
RA	Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,		
RA	Thomson N.R., James K.D., Quail M.A., Kleser H.,		
RA	Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,		
RA	Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,		
RA	Huang C.-H., Kieser T., Larke L., Murphy L.D., Oliver K., O'Neil S.,		
RA	Rabinowitz E., Rajandream M.A., Rutherford K.M., Rutter S.,		
RA	Seeger K., Saunders D., Sharp S., Squares R., Taylor K.,		
RA	Warren T., Wietzorrek A., Woodward J.R., Barrell B.G., Parkhill J.,		
RA	Hopwood D.A.;		
RT	"Complete genome sequence of the model actinomycete Streptomyces		
RT	coelicolor A3(2).";		
RL	Nature 417:141-147 (2002).		
DR	EMBL; AL939128; CAA20642.1; -.		
DR	PIR; T35237; T35237.		
DR	HSSP; P07986; 1EXG.		
DR	GO; GO:0004553; P:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.		
DR	GO; GO:0005975; P:carbohydrate metabolism; IEA.		
DR	InterPro; IPR001919; Bac celose-bind.		
DR	InterPro; IPR008965; Cellul bind.		
DR	InterPro; IPR002860; Glyco hydro_BNR.		
DR	InterPro; IPR011040; Sialidase.		
DR	Pfam; PF02012; BNR; 10.		
DR	Pfam; PF00553; CBM_2; 1.		
DR	SMART; SM00637; CBD_II; 1.		
KW	Complete proteome.		
SQ	SEQUENCE 890 AA; 93252 MW; D2C58695A4B5E84 CRC64;		
Query Match	47.4%; Score 2436; DB 2; Length 890;		
Best Local Similarity	48.6%; Pred. No. 1e-94;		
Matches	475; Conservative 129; Mismatches 248; Indels 126; Gaps 19;		
QY	11	MRSRLVLLAATAGFAVAALGVLPITATGSPA--HAATTQPYTWSNVAIGGGFVDGI	68
DB	1	MRTILITVLLAALAGLLA--GSPFAASAEPAPAAVAADSYTKNARIDGGFVPGI	57
QY	69	VNEGAPGILYVRTDIGMYRWDAANGRWIPLLDWGNNMGNGYGVWIAADPINTNKVW	128
DB	58	VFNRSKLNLAARTDIGAYRWAESSKWTPLPLDSVGSWDGHTGCVSLASDADPDRVY	117
QY	129	AAVGMNTNSWDPNGAGAILRSSDQAGATWQITPLPFKLGGMGPGRGMERLAVDPNNNDILY	188
DB	118	AAVGYTNTNSWDPNGAVLRSRDRGASQKTDLPFLKLGGMGPGRGMERLAVDPNNNDILY	177
QY	189	FGAPSGKLGWSTDSGATWSQWNPDPVGTIYANPTDITGVQSDIQGVVWVAFDKSS-SS	247
DB	178	LGAPSGHGLWRTDAGTWSEVTAEPNPGYADPNDTSGYASDNQGITWTFDESTGGG	237
QY	248	LQOASKTIFVGVADPNPNVFWSRDGAATWQAVPGAPTGFPHKGVDPDPVNVHLYIATSNIT	307
DB	238	AGTATRTLVGVADKENAVYRSTDAATWRLAGQPTGYLAHKGVLDAAENGILYLAYS	297
QY	308	GPYDSSGDVWKFVSTGTRISPVSTDTANDYFGYSGLTIDRQHPNTIMVATQISW	367
DB	298	GPYDGGKQLRYATATGTWTDISPAAEADT---YGFSGLTVDQRHPTGMATAYSSW	354
QY	368	WPDITIFRSTDGATWTRTMDWTSYPNRSIRYVLDISAEPLWTFGVQPNPPVPSPKLGWM	427
DB	355	WPDITIFRSTDGATWSQAWSYTSYPSRNRFTMDVSSSGFWLTWGANPAPPEQTPLKGMW	414
QY	428	DEAMALDPFNSDRMLYGTGATLYATNDLTKW--DSGGQIHIAPMVKGLEETAVNDLISPPS	486
DB	415	TEALEIDPFDSARMYGTGATYGTUNLWNDSGSQFTIKPMARGLEETAVNDLASPPS	474
QY	487	GAPLISALGDLGGFTHADYVAVPSTIFTSPVFTTGTSDVYAEINLPSIIVRAGSFDPSQPN	546
DB	475	GAPLISALGDLGGFRTSLTEVPSMMYTSNPNFTSTSLDPAETKPDVYVVRAGNLD--SGP	532



Qy	547	NDRHAFSTDGKKNWFQSGSEGGVTTGTTVAASADGSRFVWAPDGPQGVVYVAVFGNSW	606
Db	533	---HIAFSTDNGANWFGTDFSGVSGGTVAAGADGSRFVWSP--EGAGVQYTTGFGTSW	587
Qy	607	AASQGVPAQAQRDRVNPNTFYALSNFTFRSTGDTGVTPOVAA-GLPSSGAVGVMFHA	665
Db	588	QASTGLPAGAIIVSDRVNPFATFYGKSRFVYSTDGGATFTASAATGLPAGD--GVREKA	645
Qy	666	VPKEGDLAASS-----GLYHSTNGSSNSAITGVSSAVNVGFGKSAPSSYPAVFVV	720
Db	646	LPGEGDWMLAGGAADGYPGLWHSTDDGGTTRUPGVDAADTVGFGKAAPGASVQTLFTS	705
Qy	721	GTIGCVTCAYRSDDCGTTWVLINDDOHQYGNWGOAITGDHANLRVYIGTNGRGVYVGD1	780
Db	706	AEIGVRCIFRSTAGATWTRVNDDAHQWGTGAITGDPVYGVYVATNGRGIYGD-	764
Qy	781	GCAPSGSPSPVSPASPSLSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP	840
Db	765	-----TSDTGGTDPGPGDPTP-----	782
Qy	841	SASPSPPSP	900
Db	783	-----TGACEVYTVTNQWPGGFC--ADVRLTNTGT	811
Qy	901	S-----SVDLSS-----TVTVRYWFTF--DGGSTLVYNCW-----AAIGCGNIRA---	939
Db	812	SAWNGWSLDWSFPFGGQGVTRMNAEHTQAGTSVTRARNVGNAGVAPGASVGFPGTGSRS	871
Qy	940	-----SFGSVNPATPTTA 951	
Da	872	TNABPEGFVAVACRACPTA 889	
RESULT 3			
Qy	Q9AQH0	PRELIMINARY; PRT; 996 AA.	
AC	Q9AQH0		
DT	01-JUN-2001 (TREMBlrel. 17, Created)		
DT	01-JUN-2001 (TREMBlrel. 17, Last sequence update)		
DT	01-OCT-2003 (TREMBlrel. 25, Last annotation update)		
DE	Glycosyl hydrolase 5 (Fragment)		
OS	Caldicellulosiruptor sp. Tok7B.1.		
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;		
OC	Caldicellulosiruptor.		
OX	NCBI_TaxID=80339;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
EX	MEDLINE=20171169; PubMed=10706665;		
RA	Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,		
RA	Bergquist P.L.;		
RT	"Multidomain and multifunctional glycosyl hydrolases from the extreme		
RT	thermophile Caldicellulosiruptor isolate Tok7B.1.";		
RL	Curr. Microbiol. 40:333-340(2000).		
DR	EMBL; AF078038; AAK06388.1; -.		
DR	HSSP; Q06851; INBC.		
DR	GO; GO:0016787; P:hydrolase activity; IEA.		
DR	GO; GO:0005975; P:carbohydrate metabolism; IEA.		
DR	InterPro; IPR001956; CBD 3.		
DR	InterPro; IPR008965; Cellul. bind.		
DR	InterPro; IPR002860; Glyco_Hydro_BNR.		
DR	InterPro; IPR011040; Sialidase.		
DR	Pfam; PF02012; BNR; 9.		
DR	Pfam; PF00942; CBM 3; 1.		
DR	ProDom; PD001947; CBD 3; 1.		
KW	Hydrolase.		
FT	NON_TER		
SQ	SEQUENCE 996 AA; 108275 MW; 3C72B6ED22F3C614 CRC64;		
Query Match			
Best Local Similarity 47.1%; Score 2420.5; DB 2; Length 996;			
Matches 462; Conservative 151; Mismatches 304; Indels 35; Gaps 15;			
Qy	14	RLVSLLAATASFAVAALGVLPATITASPAHAAT---TQPYTWSNVAI-GGGGFDGIV	69

RESULT 4

Qy	Q97KK0	PRELIMINARY; PRT; 839 AA.
ID	Q97KK0	
AC	Q97KK0	
DT	01-OCT-2001 (TREMBlrel. 18, Created)	
DT	01-OCT-2001 (TREMBlrel. 18, Last sequence update)	
DT	01-MAR-2004 (TREMBlrel. 26, Last annotation update)	
DE	Probably secreted sialidase; several Asp-boxes and dockerin domain.	



```
GN OrderedLocusNames=CAC0919;
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11465286;
RA DOI=10.1128/JB.183.16.4823-4838.2001;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Ghu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L.A., Soucaille P.,
RA Daly M.J., Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838 (2001).
DR EMBL; AE007608; AAK78895.1; -.
DR PIR; D97013; D97013.
DR HSSP; P36686; 1DAV.
DR GO; GO:000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR002105; Dockerin_1.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR002860; Glyco hydro_BNR.
DR InterPro; IPR011041; Quino gluc_DH.
DR InterPro; IPR011040; Sialidase.
DR Pfam; PF02012; BNR; 9.
DR Pfam; PF00404; Dockerin_1; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_2.
KW Complete proteome.
SQ SEQUENCE 839 AA; 90824 MW; 6C99A041CA9CF984 CRC64;

Query Match 39.7%; Score 2037; DB 2; Length 839;
Best Local Similarity 49.5%; Pred. No. 5.7e-78;
Matches 392; Conservative 118; Mismatches 252; Indels 30; Gaps 16;

QY 13 SRRVLSLAA-----TASFAVAALGVLPPIAITASPAHAAT--TOPYTSNVAIGGGFVD 66
DB 2 NKRIVSMVAGLSIIPTTG-----VTHISAAANKAAASVQSQYKWNDAKIGAGGVP 54
QY 67 GIVNEGAPGLIVRYRTDTCGMYRMDAANGRIPIPLDDWGNWNGYGVVSYAADPINTNK 126
DB 55 AVIFNKTEKOLIARTDMGGAYRWDAKANKWIPITD--GFSWTMLGCESIAIDPIDNR 112
QY 127 VAAAVGMYTNSWDNDGAILRSSDQATQWITPPLPKLGGMMPGRGMRERLAVDNNNI 186
DB 113 VYIAAGLYTDQDENAYILSSQDKGNTWKRYQLPFPKVGNGMNGRGERLQIDPNDKI 172
QY 187 LYFGAPSGKLWRSTDGATWSQMTNFPDVGTYIANPTDTTGYQSDIOGVVWVAFDKSS 246
DB 173 LYLGAARSGNGLWKSSEDYGGTWSKVDNFPDPTGDYQDPQNE--YTADKVGWVWETFDPSGT 230
QY 247 SILGQAKTIFVGVAD--PNPFWSRDGGATQWAVPGAPTGFPHKGVFDPVNHVLIYATS 305
DB 231 TKGSPQTQMYGAADKTGNNIYVNDGKTSWAVKQPKGYLPHGIL--ASDGLYISVS 289
QY 306 NTGGYDGGSGDWKFSVTSQWTRISPVSTDTANDYFGYSGLTIDRQHPNTIMVATQI 365
DB 290 NTGPGYDGGQVWKYNTKTEWTNITPPAVGDTKS---GFGISVDAQNPNNVVVATLN 346
QY 366 SWPPTIIFRSTDGATWTRIDWTSYPNRSIRYVLDISAPBLTFG--VQPNPPVPSKL 424
DB 347 RWPDEEYRSTDAKTKWPIWMNGYPNRTLYGNLDYSAQWLDWGTGVTVPDPLVKL 406
QY 425 GWMDAMAI DPNSRMLYGTGATLYATNDLTWKDSGGQIHTAPMVKLEETAVNDLISP 484
DB 407 GMMGDLEIDPNSRMLYGTGATLYGTDDLLTNWDKGNVLSVXANGIEECVNDVVVP 466
QY 485 PSGAPLISALDGLGFTHADVTAVPSITFTSPVFTTGTSDVYAEINPSIIVRAGSFPSS 544
DB 467 TKGAQLLSAVGDDCCGYHDDITKVP SKMTTPNFSAITSIDYAESVPEFVVRGVNDTSK 526
QY 545 QPNDRHAFVSTDDGGKNWFQ--GSEPGGVTTGGTVAASADGSRFVWAPGDPGQPVVAVGFG 603
```

---

```
Db 527 NOQDKDCGISYDGGKWFSAAGNISGVYKAGTVAAGADAKTIVWSP--EEGANAAYSTDNG 585
QY 604 NSWAASQGVPAQAQIRSDRVNPKTFYALSNNGTFYRSTDDGGVTF-QPVAAGLSSGAVGM 662
Db 586 NKWTFCSGLPQAKVRSRVNPKPKFYFLNGKFIYSTDAGATFTOSSQTGLPTKKG-GI- 643
QY 663 FHAVPGKEGDLWLA--ASSGLVHSTNGSGSWSAITGVSSAVNVNFGKSAPESSYPVAVFVG 721
Db 644 FKTVIGHEGDIWIAGKDGDLWHSTDSGATFTKVGSDASDTVGLGSKTKDDGGYPAIYMDA 703
QY 722 TIGGVTGAYRSDCCTTWWLINDDQHYQGNWGOAITGDHANLRVRYVINGRIVYGDIG 781
Db 704 TIDGTAGIFRSDDGATWVRINDDAHQYSGPDYCICTGDPNKYGRVFGVNGRIGVYGDID 763
QY 782 GAPSGSPSPSVS 793
Db 764 GS-QPTPTPSVT 774

RESULT 5
Q70DK5 PRELIMINARY; PRT; 842 AA.
ID Q70DK5; AC Q70DK5; 05-JUL-2004 (TremBLrel. 27, Created)
DT 05-JUL-2004 (TremBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TremBLrel. 27, Last annotation update)
DE Beta-1,4-xylglucan hydrolase.
GN Name=xghA;
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN RP SEQUENCE FROM N.A.
RC STRAIN=F7;
RA Zverlov V.V.;
RL Submitted (OCN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ585344; CAB51306.1; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0000273; P:polysaccharide catabolism; IEA.
DR InterPro; IPR002105; Dockerin_1.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR002860; Glyco_hydro_BNR.
DR InterPro; IPR011040; Sialidase.
DR Pfam; PF02012; BNR; 9.
DR PROSITE; PS00448; CLOS_CELLULOsome_RPT; 2.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_2.
KW Hydrolase.
SQ SEQUENCE 842 AA; 92393 MW; CC686932DBA95FD3 CRC64;

Query Match 39.1%; Score 2007; DB 2; Length 842;
Best Local Similarity 50.8%; Pred. No. 1e-76;
Matches 399; Conservative 107; Mismatches 250; Indels 30; Gaps 16;

QY 14 RRLVLLAATASFAVAALGVLPPIAITASPAHAATTPQYTSNVAI--GGGFGVDGIVFNE 72
DB 3 KKFTSKIKAAVFAAFAVAATAIFGPAIS---SQAVTSVPYKNDVNVIGGGGFGPGIVFNE 59
QY 73 GAPGILYVRYTDIGMYRMDAANGRIPIPLDDWGNWNGYGVVSYAADPINTNKVAAV 132
DB 60 TEKDLIYARADIGAYRWDPSTETWIPLLDHFQMDYSYGVESIAIDPVDNRVIVAG 119
QY 133 MYTNSWDNDGAILRSSDQATQWITPPLPKLGGMMPGRGMRERLAVDPNNDILYFAP 192
DB 120 MYTNDWLPNWGAILRSTDRGETWETKILFPKGMNGRSMGERLAIDPNDNRILYLGR 179
QY 193 SKGLWRSTDGATWSQMTNFPDVGTYIANPTDTTGYQSDIOGVVWVAFDKSSSLGQAS 252
DB 180 CGNGLWRSTDGYGVTWSKVESFPNPGTYIYDP--NFDTYTKDIIIGVWVVFDDKSSSTPGNPT 237
QY 253 KTFIVGVADPNPFWSRDGGATQWAVPGAPTGFPHKGVFDPVNHVLIYATNTGCGPYD 312
```



Db	238	KTIVGVADKNESIYRSTGGVTKWAVQPKGLLPHHGVL-ASNGMLYITYGDTGCGPYD	296
Qy	313	GS-SDGVWKFSTGWTTRISPVSPSTDTANDYFGYSGLTIDRQHPNTIMVATQISWPDPT	371
Db	297	GNKGQVWKFNTRTGEWIDITPIPYSSDN-RFCFAGLAVDRQNPDIIMVTSMNAWPDE	355
Qy	372	IIFRSTGGATWTRIDWTSVFNPSLRVYVLDISAEPMWLTFGVQPNPPVPSPKLGWDEAM	431
Db	356	YIFRSTGGATWKNIEWGMYPERILHYEIDISAAPWLDWGTGKQLPEINPKLGMWIGDI	415
Qy	432	AIDPNSDRMLYGTGATLYATNDLTKWDSGGQIHIAPMVKGLEETAVNDLISPPSGAPLI	491
Db	416	EIDPNSDRMYVTGATTIYGCNDLTDWRGKVKIEVKATGIECAVLDLVSPPGAPLV	475
Qy	492	SALGDLGGFTHADVTAVPSTIFTSPVFTTGSVDYAEALNP-----SIIVRAGSFPDPSSQND	548
Db	476	SAVGDLVGFVHDDLKVGPKMHV-PSYSGTGIDYAEIAPFMALVAKADLYDV-----	528
Qy	549	RHVAFTDGGKNWFQ-GSEPGGVTTGGTVAAADGSRFWAPGDPQGPVVYAVFGNSWA	607
Db	529	KKISFYDGGRNWFQPPNEAPNSVGGGSVAVAADAKSVIWTB-ENASPAV-TTNDGNSWK	586
Qy	608	ASQGVPAANAIRSDRVNPKTFYALSNGTFYRSTGGVTFQPV-AAAGLPSGAVGMFHAV	666
Db	587	VCTNLGMCAVVASDRVNGKFFAFYNGKFIYSTDGLTFTDTKAPQLPKSVN---KIKAV	643
Qy	667	PKGEGDLMAA-SGLYHSTNGSSWSAITGVSSAVNNGFGKSPGSSYPAVVFVGTGG	725
Db	644	PKGEGHWLAAREGGLWRSTDGGYTFEKLNSVDTAHVVGFGKAAPQDYMAYITGKIDN	703
Qy	726	VTCAYRSDCGTTWVLINDDHOQYGNWQAITGDHANLRVYVIGTNGRGIVYGDIGGAPS	785
Db	704	VLGFFRSDAGKTWVRINDDEHGYGAVDTAITGDRPVYGRVYIATNGRGIVGYE---PAS	760
Qy	786	GSFSPS 791	
Db	761	DEPVPT 766	
RESULT 6			
ID	Q8TFP1	PRELIMINARY; PRT; 857 AA.	
AC	Q8TFP1		
DT	01-JUN-2002 (TReMBLrel. 21, Created)		
DT	01-JUN-2002 (TReMBLrel. 21, Last sequence update)		
DT	01-JUN-2003 (TReMBLrel. 24, Last annotation update)		
DE	Endoglucanase C.		
GN	Name=eglic;		
OS	Aspergillus niger.		
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;		
OC	Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.		
OX	NCBI_TaxID=5061;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=21914059; PubMed=11916668;		
RX	DOI=10.1128/AEM.68.4.1556-1560.2002;		
RA	Hasper A.A., Dekkers E., van Mil M., van de Vondervoort P.J.I.,		
RA	de Graaff L.H.;		
RT	"Eglic, a new endoglucanase from Aspergillus niger with major activity		
RT	towards xyloglucan.";		
RL	Appl. Environ. Microbiol. 68:1556-1560 (2002).		
DR	EMBL; AY040839; AAK7727.1;		
DR	GO; GO:0005576; C:extracellular; IEA.		
DR	GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.		
DR	GO; GO:0005975; P:carbohydrate metabolism; IEA.		
DR	InterPro; IPR000254; CBD_fungal.		
DR	InterPro; IPR002860; Glyco_hydro_BNR.		
DR	InterPro; IPR011040; Sialidase.		
DR	Pfam; PF02012; BNR; 7.		
DR	Pfam; PF00734; CBM_1; 1.		
DR	ProDom; PD001821; CBD_fungal; 1.		
DR	SMART; SM00236; fCBD; 1.		
DR	PROSITE; PS00562; CBD_FUNGAL; 1.		

SQ	SEQUENCE	857 AA; 90459 MW; 4822535ECD919C511 CRC64;
Query Match 35.7%; Score 1831.5; DB 2; Length 857;		
Best Local Similarity 44.4%; Pred. No. 2.5e-69;		
Matches 382; Conservative 126; Mismatches 306; Indels 47; Gaps 18;		
Qy	45	HAATQPYTWSNVAI-GGGGFDVGIIVNPGAPILYVRTDIGMYRWDANGRWIPLDW 103
Db	18	HAASQAYTKWVVTGGGGFTPGIVFNPSAKGAVARTDIGGAYRLN-SDDTWTPLMDW 76
Qy	104	VGWNWGVNGVVSIAADPINTKNVAAVGMVYNSWDPNCGAILRSSDQAGATWQITPLPEK 163
Db	77	ANNSNHWMDGIDAIATDPVDTDRVYVAVGMVYNDNDPNDGSIIRSTDQDGTWEETKLPFK 136
Qy	164	LGSNMPGRGMRLAVDPNNNDILYFGAPSGKGLWRSTDSGATWSOMTFNPDVGTIYANP 223
Db	137	VGGNMPGRGVGRLARVDPNDNSILYFGARSGNGLMKSTDYGETWSNVAFAKWTGYFQDS 196
Qy	224	TDTTGYQSDIQGVVWYAFDKSSSLGQASKTIFVGVADPNPNPVFWSRDGGATWQAVPGAP 283
Db	197	SST--YTSDPVGLAWTFDSTSGSSGSPTRIPFVGVVDTCESVFVSEDAGETWTVSGEP 254
Qy	284	T-GFTPHKGVDPVNHVLIATSNCTGPDGSSGDVWKFVSTGWTTRISPVPSDTAND 342
Db	255	MYGFLPHKGILSPSEHTLYISYNGAGPYDGTNGTVHKYNTISGVWTDISPTSMTDT--- 311
Qy	343	YFGYSLTITDRQHPNTIMVATQISWMPDTIIPSTDGGATWTRIDWTSYPNRSLRVLD 402
Db	312	YTYGGGLAVDLQVPGVTMVVAALNCWMPDELIMRSTDSGGTWSPIWANGTPSINYYSYD 371
Qy	403	ISAEPMWLTGVPQNP-PVPSPKLGMWDEAMAIDPFNSDRMLYGTGATLYATNDLTKWDSG 461
Db	372	ISNAPWLQDDTSTDEFPV--RVGMVVEALAIIDPFSDHLYGTGETIYGGHDLQNDSE 428
Qy	462	GQTHAPMVKLEETEAVNDLISPPGAPLISALGLGFTHADVTAVPSTIFTSPVFTTG 521
Db	429	HNVITSLAVGIEEMAVLGLITPPGGPALLSAVGDDGCFYHTSLTAPSQYHTPTYSST 488
Qy	522	TSVDYAEALNPSIIVRAGSFPDPSSQPNDRHVAFTDGGKNWFGQSEPGVTTGGTVAAASAD 581
Db	489	NGIDYAGNKPANIVRSGSD-----SDPTLALSSSGESWYADYAASSSTATQVALSAD 543
Qy	582	GSRFVWAPGDPQGPVVVAVGVFGNSWAAQVGPANAOIRSDRVNPKTFYALSNGTFYRSTD 641
Db	544	ADTILMNSDG-----AYRANSATLSAVSLLPSGAVIASDKANNITYFIGASGSSFYLSDD 599
Qy	642	GGVTFQVAAAGLPSSGAVGMFHAVPKEGDMLAASSGLYHSTNGSSWSAI-TGVSSA 700
Db	600	TAATF-TVITTLGSSTTANAI-RAQPSLAGDVWVSTDTGLFHSHTNYGKSFQIGSGCTEG 657
Qy	701	VNVGFGKSPGSSYPVAVVVTGGTTCAYRSDCGTTWVLINDDHOQYGN-WQAITGD 759
Db	658	WSFGFGKSPSDGDYPLFGFTVDGVTGLFKETDQGVNWQIISDAEHGFGSASANNVNGD 717
Qy	760	HANLRVYITGNGRGIVYGDIGGAPSGSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSS 819
Db	718	LQNYGVFVGTNGRGIFYGD-----PSGT-LPSATATASSASTAVKSTSTSTSTSKVGST 772
Qy	820	PSSSPSS 869
Db	773	TVSSSTATTITTSIKSTLTITTTKSSSTTSTSTSTSTSTSTSTSTSTSTSTSTSTST 831
Qy	870	SGGVKVOYKNN-----DSAPG 885
Db	832	SGWTCYENEYYSQCKSIPIG 851
RESULT 7		
ID	Q9P4T8	PRELIMINARY; PRT; 806 AA.
AC	Q9P4T8	
DT	01-OCT-2000 (TReMBLrel. 15, Created)	
DT	01-OCT-2000 (TReMBLrel. 15, Last sequence update)	



01-JUN-2003 (TReMBLrel. 24, Last annotation update)  
 DE CEL6 protein precursor.  
 GN Name=cel6;  
 OS Agaricus bisporus (Common mushroom).  
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;  
 AG Agaricales; Agaricaceae; Agaricus.  
 OC NCBI\_TaxID=5341;  
 RN [1]  
 SEQUENCE FROM N.A.  
 RC STRAIN=D649;  
 RL Morales-Almora P., Thureston C.P.;  
 Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ292929; CAC02964.1; --  
 DR GO; GO:0005576; C:extracellular; IEA.  
 DR GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . .; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR InterPro; IPR000254; CBD fungal.  
 DR InterPro; IPR002860; Glyco\_hydro\_BNR.  
 DR InterPro; IPR011040; Sialidase.  
 DR Pfam; PF02012; BNR; 8.  
 DR Pfam; PF00734; CBM\_1; 1.  
 DR SMART; SM00236; fCBD; 1.  
 DR PROSITE; PS00562; CBD\_FUNGAL; 1.  
 KW Signal.  
 FT SIGNAL. 1 23 Potential.  
 FT CHAIN 24 806 CEL6.  
 SEQUENCE 806 AA; 84777 MW; 7C6E7E382B62FF341 CRC64;  
 Query Match 34.7%; Score 1784; DB 2; Length 806;  
 Best Local Similarity 43.7%; Pred.No.2.3e-67;  
 Matches 355; Conservative 131; Mismatches 277; Indels 50; Gaps 13;  
 QY 20 LAATASFAVAALGVLPITAITASPAHAATTOPYTWSNVAI-GGGGFDVGVFNEGAPGIL 78  
 DB 1 MPSRVKFWITLTGTFSLVY-----GQASQSYRQNVKIGGGGFGVFPNSEKGLA 55  
 QY 79 YVRTDIGMYRWDANGRWIPLLDWGNNNGYVSVIAADPINTKNWAAVGMVYNSW 138  
 DB 56 YARTDIGGAYKLN-ADDTWTPLLDPADDSRWYNGVDALATDPVQPNRLYLATGLYNSW 114  
 QY 139 DPNDCAILRSSDQATWQITPLPKLGNMPCRGGERLAVDPNNDLILFEGAPSGKGLW 198  
 DB 115 DPSNGHILISSDYKGTFAAPLPPKLGNNMPCRGGERLAVDPNNLSILYFGARSGNGLW 174  
 QY 199 RSTDGGATWSOMTNPDPDVTGTIANPTDTGTQSDIQGVVWVAFDPKSSSLGQASKTIFVG 258  
 DB 175 KSTNSGRWSKVTSTDTGSGFVDPDSSTGLNSDKIGIAWITDKASGSSSATPRIFVG 234  
 QY 259 VADP--NNPVMFSRGGGATWQAVPGAPTGFIPHKGVDPDNNHVLVIATSNTPGGPYDGS 317  
 DB 235 VANKGSNSIYISTNGSSWSAVAGQPTSFLEPHKGVLSRERALVITYSDGGPYDGTSGA 294  
 QY 318 VWKFSVTGWTTRISPVSTDTANDYFGYGLTITDRQHPNTIMVATQISWPDITIIIFRST 377  
 DB 295 VTKYSIDTGAWTDITPVSGSDI---FFFGGLALDTQKNGTVMVAALNSWMPDQGIIFRST 351  
 QY 378 DGGATWTRIDWTSVPNRSRLRVLDISAEPLWITFCVQNPVPVSP---KLGWDEMAAID 434  
 DB 352 NGGASVETPLDMWAVPTLUNKYISYNAALAPW----IGNIVDVTPGNLQIGWMMESLID 407  
 QY 435 PFNSDRMLYGTGATLYATNDLTWKDSGGQIHIAPMVKGLEETAVNDLISPSPGAPLISAL 494  
 DB 408 PDSNHWLYGTGETIYGRDLLKWDASHNVTKSLADGVEETSQVALISPSPGAPLVSAL 467  
 QY 495 GDLGFTHADVTVAVPSTITPTSPVFTTGTSVDYAEINPBIIVRAGSPDPSSQPNDRHVAFS 554  
 DB 468 GDVGGFVHQSLTRAPSAQFTNPKWTSADIDFAGNPNPNVVRIGTGDS-----GKQVAIS 524  
 QY 555 TDGKKNWFGGSEPGVYTGTTVAASADGSRFFWAPDGPQPVVAVFGNSWAASQGVPA 614  
 DB 525 SDYGVTVNQHFQADPNVQGGKVAISADADITILWRTNGNG----- 563  
 QY 615 NAOIRSDKNPKTFYALNGTFYRSTDCGVTFOPVAAGLPSGSAVGVMFHAVPGKEGDLW 674



```
Db 298 SGVWTDISP---TSLASYGYGGLVDLPVQGLMVAALNCWPEDELIFRSTDSGATWS 354
QY 385 RIWDTSPNRLRYLDISAEPMLTFVQGNP-PVPSPKLGMWDEMAIDPFNSDRMLY 443
Db 355 PIWENGYPSINYYSYDISNAPMTQTTSDQFPV---RVGMVVEALAIIDPFDSNHWLY 411
QY 444 GTGATLYATNDLTWDSGGQIHAPWKVGLBETAVNDLISPSGAPLISALDGLGFTHA 503
Db 412 GTGLTVYGGHDLTNWDSKHNVTVKSLAVGIBEMAVLGLITPPGGPALLUSAVGDDGGFYHS 471
QY 504 DVTAVPSTFTSPVFTTGSTVDYAEINPSIIRAGSFPDPSOPNDRHVAFTSDGCKWQF 563
Db 472 DIDAAPNOAYTPPYGTNGIDYAGNKPENIVRSGAD--DYPT---LALSNFSGTWYA 526
QY 564 GSEPGVTTGGTVAASADSRFPWAPDGPQPVVYAVFGNSWAASQGVPAQAIRSDRV 623
Db 527 DYAASTSTGTGAVALSADGDTVLLMSSTSGALVSKSQG---TLTAVSSLPSCGAVIASDKS 583
QY 624 NPKTYALNNGTFYRSTGGTTFQPVAAGLPSSGAVGVWFHAPVKEGDLMLAASGLYH 683
Db 584 DNTVYFGSAGALYVSKNTATSTFTKVS--LGSSTTVNAI--RAHPSIAGDVWASTDKGLWH 641
QY 684 STNGSSWSAI--TGVSANVNGFGKSPGSSYPVAVVVGTTIGVTVGAYRSDDCGTTWVLI 742
Db 642 STDYGSTFTQIGSGVTAGSFGKASSTGVIIYGFPTIDGAAGLKFSEADAGTNQVLI 701
QY 743 NDDQHQYNWQA--ITGDHANLRVYIGTNGRGIVYGDIGGAPSG-----SPSPSV 792
Db 702 SDASHFGS--GSANVNGDLQYGRVFRGHRPGLLRQSQREPAGRHGDDGDDTTSTKT 760
QY 793 SPASPSLSPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 852
Db 761 STTVSTLTKTTSSASTSSSTTVKTTTSSSTTSKASSTTTTSTTTSSGTTATA 820
QY 853 SS 854
Db 821 SA 822

RESULT 9
Q729M8 PRELIMINARY; PRT; 838 AA.
AC Q729M8;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Cel74a.
OS Trichoderma reesei (Hypocrea jecorina).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
OX NCBI_TaxID=51453;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=QM6a;
RA Foreman P.K., Brown D., Dankmeyer L., Dean R., Diener S.,
RA Dunn-Coleman N.S., Goedegebuur P., Houfek T.D., England G.J.,
RA Kelley A.S., Meerman H.J., Mitchell T., Mitchinson C., Olivares H.A.,
RA Teunisissen P.J.M., Yao J.J., Ward M.;
RA "Transcriptional Regulation of Biomass-Degrading Enzymes in the
RT Filamentous Fungus Trichoderma reesei."
RL J. Biol. Chem. 0:0-0(2003).
RL EMBL: AY281371; RAS7752.1; -.
DR GO: 0005576; C:extracellular; IEA.
DR GO: 0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO: 0005975; P:carbohydrate metabolism; IEA.
DR InterPro: IPR000254; CBD_fungal.
DR InterPro: IPR002860; Glyco hydro_BNR.
DR InterPro: IPR011040; Sialidase.
DR Pfam: PF02012; BNR; 8.
DR Pfam: PF00734; CBM_1; 1.
DR ProDom: PD001821; CBD_fungal; 1.
```

```
DR SMART; SM00236; fCBD; 1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
SQ SEQUENCE 838 AA; 87132 MW; FBE3D08F2A74FA84 CRC64;

Query Match 31.7%; Score 1625.5; DB 2; Length 838;
Beat Local Similarity 41.7%; Pred. No. 1.1e-60;
Matches 350; Conservative 131; Mismatches 296; Indels 63; Gaps 23;

QY 25 SPFAVAALGLVPIATASPAHAATTPQYTSNVAI--GGGGFVDGIVFNEGAPGILYVRTD 83
Db 4 SRLVALVLGV-----IPAHAA-----FSWKVKVKGGGGFGVGIIFHPKTKGVAYARTD 53
QY 84 ICGMYRWDAANGRWIPLLDWV----GWNWNGVGVVSIADPINTNKVMAAVGMYTNSWD 139
Db 54 IGLYRLN--ADDSWTAVTDDGIADNAGHNW---GIDAVALDPQDDQKVAAVAGMYTNSWD 109
QY 140 PNDGAILRSSDGCATWITPLPKLGGNMPGRGMBERLAVDPNDNLLYFGAPSGKGLWR 199
Db 110 PENGAIIRSSDRGATWSFTNLFPKVGNNMPGRGAGERLAVDPANSNIIFYGARSNGLWK 169
QY 200 STDGATWSQMTNFPDVGTYIANPTDITGVQSDIQGVVWVAFDKSSSSILGOASKTIIFGV 259
Db 170 STDGGVTFSKVSFTATGYIIPDPSDNGSNYDKQGLMMVTFDSTSTTCGATSRI FVGT 229
QY 260 ADP--NNPVFWSRDGGATQAVPGAPTGFIPHKGVDFVNVHLYIATSNTPGPDSSGDV 318
Db 230 ANNIATSVVSTNAGSTWSAVFGQPKYFPHKAKLQPAEKALYLYTSDGTGPDGTLGSV 289
QY 319 WKFSVTSGTWTRISVPSTDTANDYFGYSLGTLTIDRQHPNTIMVATQISWPDITIIFRSTD 378
Db 290 WRDYIAGGTWKDITPVSGSDL---YFGFGGLGLDLQKPGTLVVASLNSWMPDAQLFRSTD 346
QY 379 GGATWTRIDWTSYPNRSLRYVLDISAEKWLTFG--VQNPFPVPS---PKLGMWDEMAAI 433
Db 347 SGTWSPPIAWASYPTEYTYISISTPKAPWIKNFIIDVTSESDDGLIKLGMWDEMAI 406
QY 434 DPFNSDRMLYGTGATLYATNDLTWDSGGQIHIAPMKVGLBETAVNDLISPPSGAPLISA 493
Db 407 DPTDSNHWLYGTGWTIFGGHDLTNWDRHNSIQSLADGIEEFVQDLASAPGSELLAA 466
QY 494 LGDLGGLGTHA---DVTAVPSTFTSPVFTTGSTVDYAEINPSIIRAGSFPDPSOPNDRH 550
Db 467 VGDDNGFTFASRNDLGTSPQVWATPTWATSTSDYAGNSVKSVMVVGNTAGTQQ----- 521
QY 551 VAFSTDGKKNWFOGSEPGVTTGCTVAASADGRFWAPDGPQPVVYAVFGNSWAASQ 610
Db 522 VAISDGGATWSIDYAADTSMNGTIVAYSADGPTILWSTASSG---VQRSQFQGSFASVS 578
QY 611 GVPANAOIBSDRVNPKTFVALSNGTFRSTGVTGQPVAAGLPSSGAVGVW---FHAVPG 668
Db 579 SLPAGAVIASDKKTNSVYFAGSGSTFVSKDTGSSP---TRG-PKLSAGTIRIDIAHPT 634
QY 669 KEGDLMLAASSGLYHSTNGSSWSAI--TGVSANVNGFGKSPGSSYPVAVVVGTTIGV 727
Db 635 TAGTLVSTDVGIIFRSTDSGTTFGQVSTALTNTYQIALGVGS--GSNW-NLVAFGT--GPS 690
QY 728 GA--YRSDDCGTTWVLIINDDQHQYGNWQOAITGDHANLRVYIGTNGRGIYV--GDIGA 783
Db 691 GARLYASGDSGASWTDIQGSQSGSIDSTKVAGSGTAGQVYVGTNGRGVFAQGTVGCG 750
QY 784 PSQSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 843
Db 751 TGGT-----SSSTKQSSSTSTSSSTTLRVSSTVSTRASVTSSRTSSAAGPTGS 801

RESULT 10
Q7S1W6 PRELIMINARY; PRT; 861 AA.
AC Q7S1W6;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein.
```



```
GN Name=NCU05955.1;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OR74A;
RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,
RA Elkins T., Engels P., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,
RA Selitrenikoff C.P., Kinsey J.A., Braun E.L., Zeller A., Schulte U.,
RA Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,
RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,
RA Kamal M., Kanvasselis M., Mauceli E., Bielke C., Rudd S., Frisman D.,
RA Krystofova S., Rasmussen C., Metzenberg R.B., Perkins D.D., Kroken S.,
RA Cogoni C., Macino G., Catcheside D., Li W., Pratt R.J., Osmari S.A.,
RA DeSouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,
RA Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
RA Natvig D.O., Alex L.A., Mannhaupt G., Ebbole D.J., Freitag M.,
RA Paulsen I., Sachs M.S., Lander E.S., Nisbaum C., Birren B.;
RT "The Genome Sequence of the Filamentous Fungus Neurospora crassa.";
RL Nature 0:0(2003).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC
DR EMBL; AAX01000439; EAA29333.1; -.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR000254; CBD_fungal.
DR InterPro; IPR002860; Glyco_hydro_BNR.
DR InterPro; IPR011041; Quino gluc DH.
DR InterPro; IPR011040; Sialidase.
DR Pfam; PF02012; BNR; 9.
DR Pfam; PF00734; CBM.1; 1.
DR ProDom; PD001821; CBD_fungal; 1.
DR PROSITE; PS00455; AMP_BINDING; UNKNOWN_1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
DR KX Hypothetical protein.
SQ SEQUENCE 861 AA; 89114 MW; 95BDB846631C67D CRC64;

Query Match 30.9%; Score 1587.5; DB 2; Length 861;
Best Local Similarity 38.0%; Pred. No. 4.5e-59;
Matches 356; Conservative 169; Mismatches 306; Indels 107; Gaps 27;

QY 11 MRSRLVSLAATASFAVAALGVLPITAITASPAHAATQPVTSNVAI-GGGGFVDGIV 69
DB 1 MKSISVQLLAAA-----AAPVRAA-----SWKNVIGGGGFVPGFV 40
QY 70 FNEGAGILYVRTDIGMYRWAANGRIWIPLLDWV-----GWNNGYGVWSIAADPINTN 125
DB 41 FHPTEKGAVARTDIGLYRLN-ADDSMTAITSITDDKNNW---GIDAVALDPQDPD 96
QY 126 KWAAVGMYTNSDPDGAIIIRSDGATWQITPLPKLGGNMPGRCGERLAVDPNRDN 185
DB 97 KYVAATGYTNSDPNNGTLIRSSDRGATWSSTELPFKVGGMPCRGGERLAVDPKNSK 156
QY 186 ILYFGAPSGKLWRSTDSGATWSQNTNFPDVGTYIANPTDITGYSQSDIQGVVWVAFKSS 245
DB 157 ILYFGARSGHLYKSTDSGVTFSKVSTFAGVYVVDPSDITGLNNDKQGFATVTFDSTS 216
QY 246 SSLGQAASKTIFVGVADPNPNPVFW-SRDGGATWQAVPGAPTGFIPHKGVDFPNVHLYIAT 304
DB 217 ATTNGATSRIFVGATNTSSVWVNSDAGATWSAVAGPGYFFPHKCKLQTERALYLYT 276
QY 305 SNTGPGYDGGSDGVKFSVTGWTTRISPVSTDTANDYFGVSLGITDRQHPNTIMVATQ 364
DB 277 SDGTGPGYDGTAGAVRVYDITNGTKDIT--PGSDLA---YGFGLGVDMKNPPTIMVASL 331
QY 365 ISWHPDITIFRSTDGGATWIRNDWTSYPNLSRLVLDISAEPLWTFGVQNPFPVPSPKL 424
```















Db	508	QFSLDSIDAAGNFPNVVRAGSSHEYSACARGAYATDGDWTFPTCPFGMNASHY	567
Qy	573	-GGTVAASADGSRFWAP--GDFGQPVVYAVGFCNSWAASQGYPA-----NAQIRSDRV	623
Db	568	QGSTIAVDASGSGQIVMSSTKLDEQASGWPYSHDYKGTWS----VPAGDLKAQTANVLSDKV	623
Qy	624	NPKTFYALSNGTFFVRSSTGGVTFQPVAAAGLPSSGAVGVMFHAVPGKEGDLWL-AASSGLY	682
Db	624	QDGTFFATDGGKFFVSTDGKSYAAKAGGLVT--GTSMLPFAVNPWVAGDVWVPVPEGGLF	681
Qy	683	HSTNGSSWSAI-TGVSSAYNVGFGKS----APGSSYPVAVVVGTT--IGGVTGAYRSDDC	735
Db	682	HSTDGASFTRVGTANATLVSGAPKSKSDGKKASAPSAVFIWGTDKPGSDIGLYRSDDN	741
Qy	736	GTTWVLINDQHQYGNWGOAITGDHANLRVYIGTNGRGIVYDI	780
Db	742	GSTWTRVNDQEHNYSG-PTWIEADPKVYGRVYLGTVGTNGRGIVYADL	785

Search completed: October 5, 2005, 07:56:03  
Job time : 124.987 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 07:49:28 ; Search time 36.7684 Seconds  
(without alignments)  
1942.949 Million cell updates/sec

Title: US-09-917-376-1

Perfect score: 5135

Sequence: 1 MDRSENRLTWRSRLVSL.....RASPGSVNPATPTADTYLQX 957

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*

- 1: /cgm2\_6/ptodata/1/iaa/5A COMB.pap.\*
- 2: /cgm2\_6/ptodata/1/iaa/5B COMB.pap.\*
- 3: /cgm2\_6/ptodata/1/iaa/6A COMB.pap.\*
- 4: /cgm2\_6/ptodata/1/iaa/6B COMB.pap.\*
- 5: /cgm2\_6/ptodata/1/iaa/PCITUS COMB.pap.\*
- 6: /cgm2\_6/ptodata/1/iaa/backfiles1.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query No.	Score	Match	Length	DB ID	Description
1	334	6.5	1751	3	US-09-136-574A-44	Sequence 44, Appl
2	301.5	5.9	1749	4	US-09-640-419C-28	Sequence 28, Appl
3	288	5.6	96	4	US-09-119-507B-112	Sequence 112, Appl
4	288	5.6	96	4	US-09-547-693-236	Sequence 236, Appl
5	276	5.4	1426	3	US-09-136-574A-43	Sequence 43, Appl
6	260	5.1	2736	4	US-09-252-991A-30227	Sequence 30227, A
7	234	4.6	2137	2	US-09-134-001C-4463	Sequence 4463, Ap
8	225.5	4.4	551	2	US-09-033-537A-1	Sequence 1, Appli
9	216.5	4.2	8991	4	US-08-714-741-32	Sequence 32, Appl
10	215.5	4.2	76	4	US-09-547-693-233	Sequence 233, App
11	213	4.1	3892	4	US-09-328-352-5503	Sequence 5503, Ap
12	210.5	4.1	521	1	US-08-276-213-3	Sequence 3, Appli
13	208.5	4.1	206	4	US-08-529-055-54	Sequence 54, Appl
14	208.5	4.1	493	3	US-09-198-956-10	Sequence 10, Appl
15	208.5	4.1	493	3	US-09-198-955A-12	Sequence 12, Appl
16	208.5	4.1	493	3	US-09-694-531-12	Sequence 12, Appl
17	208.5	4.1	493	4	US-09-670-141-10	Sequence 10, Appl
18	208.5	4.1	493	4	US-10-072-152-12	Sequence 12, Appl
19	205	4.0	918	4	US-09-200-650E-1	Sequence 1, Appli
20	204.5	4.0	1970	4	US-09-538-092-1005	Sequence 1005, Ap
21	202.5	3.9	476	4	US-09-339-159B-4	Sequence 4, Appli
22	200.5	3.9	700	2	US-07-862-588B-2	Sequence 2, Appli
23	200	3.9	183	4	US-08-529-055-50	Sequence 50, Appl
24	199.5	3.9	490	3	US-09-109-841-2	Sequence 2, Appli
25	199.5	3.9	616	3	US-09-136-574A-47	Sequence 47, Appl
26	197	3.8	1060	3	US-08-911-393-2	Sequence 2, Appli
27	197	3.8	1060	4	US-09-955-909-2	Sequence 2, Appli

28 192.5 3.7 271 4 US-09-248-796A-21050 Sequence 21050, A  
29 191.5 3.7 412 1 US-08-313-288B-18 Sequence 18, Appl  
30 191 3.7 105 4 US-09-547-693-230 Sequence 230, App  
31 191 3.7 2169 4 US-09-949-016-6930 Sequence 6930, Ap  
32 190 3.7 1092 4 US-09-147-405B-15 Sequence 15, Appl  
33 189.5 3.7 1187 4 US-09-949-016-6513 Sequence 6513, Ap  
34 189 3.7 933 3 US-08-293-728-2 Sequence 2, Appli  
35 189 3.7 933 3 US-09-421-868-2 Sequence 2, Appli  
36 188.5 3.7 193 4 US-08-956-171E-5249 Sequence 49, Appl  
37 187 3.6 936 4 US-08-529-055-49 Sequence 49, Appl  
38 187 3.6 936 4 US-08-781-986A-5249 Sequence 5249, Ap  
39 186.5 3.6 104 4 US-09-547-693-235 Sequence 235, App  
40 186 3.6 423 2 US-08-760-797A-1 Sequence 1, Appli  
41 185.5 3.6 424 3 US-08-932-929B-1 Sequence 1, Appli  
42 185.5 3.6 793 4 US-09-270-767-42801 Sequence 42801, A  
43 185.5 3.6 1719 2 US-08-459-568-4 Sequence 4, Appli  
44 185.5 3.6 1719 2 US-08-399-411-4 Sequence 4, Appli  
45 185.5 3.6 1719 3 US-08-516-859A-4 Sequence 4, Appli

#### ALIGNMENTS

#### RESULT 1

US-09-136-574A-44

; Sequence 44, Application US/09136574A

; Patent No. 6294366

; GENERAL INFORMATION:

; APPLICANT: Farrington, Graham K.

; Anderson, Paige

; Gibbs, Moreland

; Bergquist, Peter

; Daniels, Roy

; Morgan, Hugh W.

; Williams, Diane P.

; TITLE OF INVENTION: Compositions and Methods for

; Treating Cellulose Containing

; Cellulase Enzyme Compositions

; NUMBER OF SEQUENCES: 49

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Howson and Howson

; STREET: Spring House Corporate Center, P.O. Box 457

; CITY: Spring House

; STATE: PA

; COUNTRY: USA

; ZIP: 19477

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/136,574A

; FILING DATE: 19-Aug-1998

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/932,571

; FILING DATE: September 19, 1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Bak, Mary E.

; REGISTRATION NUMBER: 31,215

; REFERENCE/DOCKET NUMBER: 1997US001/CIP

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 215-540-9200

; TELEFAX: 215-540-5818

; TELEX: <Unknown>

; INFORMATION FOR SEQ ID NO: 44:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1751 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; SEQUENCE DESCRIPTION: SEQ ID NO: 44:







Db	503	-GNLTFDPTITIDALFS-----LGTGSIIPANTGPINTGNTLVTSPPTNLDNDKNIQLNA	557
Qy	467	-----AMVKGLETAVN-----DLISPPS-----GAPLISALGDLG-----	498
Db	558	NTTFTAPGYDIYFRKSVNGGFDLGNANFVYFDDGAGITTPKLSFGVTATYIYVGNDIVT	617
Qy	499	-----GFTHADVT-AVPTIFTSPVFTTG-----TSVDYAEILNP	531
Db	618	QGNQIFGVFLQPVNLTTSAGSVIFTNINLLNGSLQVTAQNVISQPSSSLGAVEIAS	677
Qy	532	SIIVRAGSFDPSSQPNDRHNAF-----STDGGRNWFQGSBPGGVTTG-----GTVA	577
Db	678	DVLLNAG-----QNVSGFNINTRGGNVDIQAL--GNISTGSIVTSFPGGNAGNVI	725
Qy	578	ASADG-----SRFVWAPGDPQPVVAVGFGNSWA-----ASQGV	612
Db	726	LNAGGTLTGVYIETSGTNGGVDVTTSSGSNTSTAVIDTRGFGDGLDLSLGGAVSIESKGD	785
Qy	613	PANAQIRSDRVNPKTFYALSNGTFRSDTGGVTFQPVAAAGLPSSGAVGVMFHAVPGKE--	670
Db	786	ITTAFTIDTGAYSIESFNGTGNVFLTAGDSITTYIPTAGKNGG--DIFFOAGESIIL	843
Qy	671	-----GDLWLA-----SSGLYHSTNGGS-----SWSAITGV	697
Db	844	DYLTNYGQTSQDGVVVEAPLDISGSYIYTGSGEPGNVFLQAGDITTSYIDTSAANGG	903
Qy	698	SSAVNVG-----FGKAPGSSYPVAVVVGT	722
Db	904	DIFIQSGGDEVGYLFTKGYEGRGDGVVETGRYFRAIDGFLGEGFPFVYTAGLTVG-	962
Qy	723	IGVVTGAYRSDDCGTTWVLINDDQHOYGNWQAITGDHANLRVYIGTNGR--GIVYGD	780
Db	963	-----GSVYI-----QFGSFPFIIGNPI-----TNGTGAISSGDD	994
Qy	781	CGAPSGSP-----SPSVSPASPSLSRSPSPSSSPSSSPSPSSSPSSSPSS	825
Db	995	NTVPITGTPPIEDFTFLDNITTEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPE	1054
Qy	826	PSPPSPSPSPSPSPSPSPSPSPSPSSSPSSSPSSSPSPSSSPSPSSSP	868
Db	1055	PE	1097
RESULT 3			
US-09-119-507B-112			
; Sequence 112, Application US/09119507B			
; Patent No. 6548642			
GENERAL INFORMATION:			
; APPLICANT: Kieliszewski, Marcia J.			
; TITLE OF INVENTION: No. 6548642e1 Synthetic Genes for Plant Gums			
; FILE REFERENCE: OHU-03417			
; CURRENT APPLICATION NUMBER: US/09/119,507B			
; CURRENT FILING DATE: 1998-07-20			
; NUMBER OF SEQ ID NOS: 118			
; SOFTWARE: PatentIn Ver. 2.0			
; SEQ ID NO 112			
; LENGTH: 96			
; TYPE: PRT			
; ORGANISM: Artificial Sequence			
; FEATURE:			
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic			
US-09-119-507B-112			
Query Match 5.6%; Score 288; DB 4; Length 96;			
Best Local Similarity 77.0%; Pred. No. 1.6e-11;			
Matches 57; Conservative 4; Mismatches 13; Indels 0; Gaps 0;			
Qy	791	SVSPSPSPSPSPSPSPSPSPSPSSSPSSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP	850
Db	20	SLSLAQTTRASP	79
Qy	851	SPSSSPSPSPSPPTP	864
TITLE OF INVENTION: Compositions and Methods for Treating Cellulose Containing Fabrics Using Truncated Cellulase Enzyme Compositions			
NUMBER OF SEQUENCES: 49			
CORRESPONDENCE ADDRESS:			
ADDRESSEE: Howson and Howson			
STREET: Spring House Corporate Center, P.O. Box 457			
CITY: Spring House			
STATE: PA			
COUNTRY: USA			
ZIP: 19477			
COMPUTER READABLE FORM:			
MEDIUM TYPE: Diskette			
COMPUTER: IBM Compatible			
OPERATING SYSTEM: DOS			
CURRENT APPLICATION DATA:			
APPLICATION NUMBER: US/09/136,574A			
FILING DATE: 19-Aug-1998			
CLASSIFICATION: <Unknown>			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: US 08/932,571			
FILING DATE: September 19, 1997			
ATTORNEY/AGENT INFORMATION:			
US-09-547-693-236			
Query Match 5.6%; Score 288; DB 4; Length 96;			
Best Local Similarity 77.0%; Pred. No. 1.6e-11;			
Matches 57; Conservative 4; Mismatches 13; Indels 0; Gaps 0;			
Qy	791	SVSPSPSPSPSPSPSPSPSPSPSSSPSSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP	850
Db	20	SLSLAQTTRASP	79
Qy	851	SPSSSPSPSPSPPTP	864
TITLE OF INVENTION: Compositions and Methods for Treating Cellulose Containing Fabrics Using Truncated Cellulase Enzyme Compositions			
NUMBER OF SEQUENCES: 49			
CORRESPONDENCE ADDRESS:			
ADDRESSEE: Howson and Howson			
STREET: Spring House Corporate Center, P.O. Box 457			
CITY: Spring House			
STATE: PA			
COUNTRY: USA			
ZIP: 19477			
COMPUTER READABLE FORM:			
MEDIUM TYPE: Diskette			
COMPUTER: IBM Compatible			
OPERATING SYSTEM: DOS			
CURRENT APPLICATION DATA:			
APPLICATION NUMBER: US/09/136,574A			
FILING DATE: 19-Aug-1998			
CLASSIFICATION: <Unknown>			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: US 08/932,571			
FILING DATE: September 19, 1997			
ATTORNEY/AGENT INFORMATION:			
US-09-547-693-236			
Query Match 5.6%; Score 288; DB 4; Length 96;			
Best Local Similarity 77.0%; Pred. No. 1.6e-11;			
Matches 57; Conservative 4; Mismatches 13; Indels 0; Gaps 0;			
Qy	791	SVSPSPSPSPSPSPSPSPSPSPSSSPSSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP	850
Db	20	SLSLAQTTRASP	79
Qy	851	SPSSSPSPSPSPPTP	864
TITLE OF INVENTION: Compositions and Methods for Treating Cellulose Containing Fabrics Using Truncated Cellulase Enzyme Compositions			
NUMBER OF SEQUENCES: 49			
CORRESPONDENCE ADDRESS:			
ADDRESSEE: Howson and Howson			
STREET: Spring House Corporate Center, P.O. Box 457			
CITY: Spring House			
STATE: PA			
COUNTRY: USA			
ZIP: 19477			
COMPUTER READABLE FORM:			
MEDIUM TYPE: Diskette			
COMPUTER: IBM Compatible			
OPERATING SYSTEM: DOS			
CURRENT APPLICATION DATA:			
APPLICATION NUMBER: US/09/136,574A			
FILING DATE: 19-Aug-1998			
CLASSIFICATION: <Unknown>			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: US 08/932,571			
FILING DATE: September 19, 1997			
ATTORNEY/AGENT INFORMATION:			
US-09-547-693-236			



























GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 07:52:23 ; Search time 121.54 Seconds  
(without alignments)  
3268.602 Million cell updates/sec

Title: US-09-917-376-1  
Perfect score: 5135  
Sequence: 1 MDRSENIRLTMSRRRLVSL...RASFGSNPATPTADTVLQX 957

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1846076 seqs, 415116000 residues  
Total number of hits satisfying chosen parameters: 1846076

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/FCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10E\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US10F\_PUBCOMB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 19: /cgn2\_6/ptodata/1/pubpaa/US11A\_PUBCOMB.pep.\*
- 20: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*
- 21: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 22: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5134	100.0	957	11	US-09-917-376-1
2	5134	100.0	957	14	US-10-155-400-1
3	4036	78.6	740	11	US-09-917-376-3
4	4036	78.6	740	11	US-09-917-376-6
5	4036	78.6	740	14	US-10-155-400-3
6	4036	78.6	740	14	US-10-155-400-6
7	2478	48.3	882	14	US-10-156-761-9395
8	1680	32.7	726	11	US-09-917-376-7
9	1680	32.7	726	14	US-10-155-400-7
10	1625.5	31.7	838	15	US-10-420-191-2
11	1604	31.2	818	14	US-10-026-994-2

12	1442	28.1	739	14	US-10-156-761-10111	Sequence 10111, A
13	1125	21.9	812	15	US-10-395-241-12	Sequence 12, Appl
14	1117	21.8	789	15	US-10-395-241-14	Sequence 14, Appl
15	1114	21.7	826	15	US-10-395-241-18	Sequence 18, Appl
16	848	16.5	555	10	US-09-927-827-47	Sequence 47, Appl
17	823	16.0	1228	10	US-09-917-384-1	Sequence 1, Appl
18	823	16.0	1228	10	US-09-917-383-1	Sequence 1, Appl
19	702	13.7	762	10	US-09-917-378-1	Sequence 1, Appl
20	511.5	10.0	1043	10	US-09-917-384-6	Sequence 6, Appl
21	511.5	10.0	1043	10	US-09-917-383-6	Sequence 6, Appl
22	462	9.0	88	11	US-09-917-376-5	Sequence 5, Appl
23	462	9.0	88	14	US-10-155-400-5	Sequence 5, Appl
24	462	9.0	89	11	US-09-917-376-4	Sequence 4, Appl
25	462	9.0	89	14	US-10-155-400-4	Sequence 4, Appl
26	462	9.0	154	10	US-09-917-378-4	Sequence 4, Appl
27	459	8.9	150	10	US-09-917-384-5	Sequence 5, Appl
28	459	8.9	150	10	US-09-917-383-5	Sequence 5, Appl
29	389	7.6	284	17	US-10-418-032-270	Sequence 270, App
30	346.5	6.7	901	17	US-10-939-262-12	Sequence 12, Appl
31	336.5	6.6	741	14	US-10-156-761-8100	Sequence 8100, Ap
32	326.5	6.4	628	17	US-10-939-262-14	Sequence 14, Appl
33	288	5.6	96	15	US-10-437-708-236	Sequence 236, App
34	288	5.6	96	15	US-10-395-402-112	Sequence 112, App
35	288	5.6	96	16	US-10-257-199-236	Sequence 236, App
36	288	5.6	96	17	US-10-418-032-236	Sequence 236, App
37	274	5.3	406	18	US-10-450-763-57609	Sequence 57609, A
38	268	5.2	284	17	US-10-418-032-271	Sequence 271, App
39	262	5.1	2468	14	US-10-248-330-4	Sequence 4, Appl
40	262	5.1	2468	15	US-10-282-122A-66335	Sequence 66335, A
41	261.5	5.1	599	10	US-09-955-555A-29	Sequence 29, Appl
42	253	4.9	2435	15	US-10-282-122A-47453	Sequence 47453, A
43	252.5	4.9	2117	15	US-10-120-801-63	Sequence 63, Appl
44	246	4.8	1066	16	US-10-647-195-43	Sequence 43, Appl
45	243.5	4.7	1049	15	US-10-282-122A-49900	Sequence 49900, A

ALIGNMENTS

RESULT 1  
US-09-917-376-1  
; Sequence 1, Application US/09917376  
; Publication No. US20040038334A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOYTICUS  
; FILE REFERENCE: 40197.4US01  
; CURRENT APPLICATION NUMBER: US/09/917,376  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 957  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; NAME/KEY: MOD RES  
; LOCATION: (957)  
; OTHER INFORMATION: Any amino acid  
US-09-917-376-1

Query Match 100.0%; Score 5134; DB 11; Length 957;  
Best Local Similarity 100.0%; Pred. No. 3.9e-280;  
Matches 956; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MDRSENIRLTMSRRRLVSLAATAAFVAAALGVLPITAITASPAHAATTPQYTSNVAIG 60  
Db 1 MDRSENIRLTMSRRRLVSLAATAAFVAAALGVLPITAITASPAHAATTPQYTSNVAIG 60



```

QY 61 GGGFVDGIVFNEGAPGILYVRTDIGMYRWDAANGRWIPLLDWYGNWNGYGVVIAAD 120
DB 61 GGGFVDGIVFNEGAPGILYVRTDIGMYRWDAANGRWIPLLDWYGNWNGYGVVIAAD 120
QY 121 PINTNKVAAVGMVYNSWDNDGAILRSSDQATWQITPLPFKLGGMNPGRMGERLAVD 180
DB 121 PINTNKVAAVGMVYNSWDNDGAILRSSDQATWQITPLPFKLGGMNPGRMGERLAVD 180
QY 181 PNNDNILYFGAPSGKGLWRSYVTDIGMYRWDAANGRWIPLLDWYGNWNGYGVVIAAD 240
DB 181 PNNDNILYFGAPSGKGLWRSYVTDIGMYRWDAANGRWIPLLDWYGNWNGYGVVIAAD 240
QY 241 FDKSSSSLGQASKTIFVGVADPNPNVFWRSRDCGATWQAVPGAPTGFIPHKGVFDPVNHVL 300
DB 241 FDKSSSSLGQASKTIFVGVADPNPNVFWRSRDCGATWQAVPGAPTGFIPHKGVFDPVNHVL 300
QY 301 YIATSNTPGPGYDSSGDKWKFSTVSGTWTRISPVSTDTANDYFGYSGLTIDRQHPNTIM 360
DB 301 YIATSNTPGPGYDSSGDKWKFSTVSGTWTRISPVSTDTANDYFGYSGLTIDRQHPNTIM 360
QY 361 VATQISWMPDTIIFRSTDCGATWTRIDWTSYPNRSRYVLDISAEPWLTFFGVQPNPVP 420
DB 361 VATQISWMPDTIIFRSTDCGATWTRIDWTSYPNRSRYVLDISAEPWLTFFGVQPNPVP 420
QY 421 SPKLGWDEAMAIIDPFNSDRMLYGTGATLYATNDLTWDSGGQIHIAPMVKGLEETA VND 480
DB 421 SPKLGWDEAMAIIDPFNSDRMLYGTGATLYATNDLTWDSGGQIHIAPMVKGLEETA VND 480
QY 481 LISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTGSVDYAEINPSIIVRAGSF 540
DB 481 LISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTGSVDYAEINPSIIVRAGSF 540
QY 541 DPSSQPNDRHVAFTSDGCKNWFQSGPGVTTGGTVAASADGSRFVWAPGPGQPVVYAV 600
DB 541 DPSSQPNDRHVAFTSDGCKNWFQSGPGVTTGGTVAASADGSRFVWAPGPGQPVVYAV 600
QY 601 GFNGSMAASQGVPAANAQIRSDRNPVFWRSRDCGATWQAVPGAPTGFIPHKGVFDPVNHVL 660
DB 601 GFNGSMAASQGVPAANAQIRSDRNPVFWRSRDCGATWQAVPGAPTGFIPHKGVFDPVNHVL 660
QY 721 GTIGGVTAAYRSDDCGTTWVINDQHOYGNWQAITGDHANLRVYIGTNGRGIVYGD 780
DB 721 GTIGGVTAAYRSDDCGTTWVINDQHOYGNWQAITGDHANLRVYIGTNGRGIVYGD 780
QY 781 GGAPSGSPSVSPSASPSSPSPPSSPSPPSSPSPPSSPSPPSSPSPPSSPSPPSSPSPP 840
DB 781 GGAPSGSPSVSPSASPSSPSPPSSPSPPSSPSPPSSPSPPSSPSPPSSPSPPSSPSPP 840
QY 841 SASPSSSSPSPPSSPSPPSSPSPPSSPSPPSSPSPPSSPSPPSSPSPPSSPSPPSS 900
DB 841 SASPSSSSPSPPSSPSPPSSPSPPSSPSPPSSPSPPSSPSPPSSPSPPSSPSPPSS 900
QY 901 SSVDLSTVTVRYWTFTRDGGSTLVYNCWAAIGCGNIRASFGSVNPAFTADTYLQ 956
DB 901 SSVDLSTVTVRYWTFTRDGGSTLVYNCWAAIGCGNIRASFGSVNPAFTADTYLQ 956

```

RESULT 2

```

US-10-155-400-1
; Sequence 1, Application US/10155400
; Publication No. US20030108988A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOYTICUS
; FILE REFERENCE: NREL 01-36A

```

```

; CURRENT APPLICATION NUMBER: US/10/155,400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 957
; TYPE: PRT
; ORGANISM: Acidothermus cellulosyticus
; NAME/KEY: MOD_RES
; LOCATION: (957)
; OTHER INFORMATION: Any amino acid
; US-10-155-400-1

Query Match 100.0%; Score 5134; DB 14; Length 957;
Best Local Similarity 100.0%; Pred. No. 3.9e-280;
Matches 956; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDRSENIRLTWRSRRLVSLAATASFAVAALGLVLPALITASPAHAATTQPYTWSNVAIG 60
DB 1 MDRSENIRLTWRSRRLVSLAATASFAVAALGLVLPALITASPAHAATTQPYTWSNVAIG 60
QY 61 GGGFVDGIVFNEGAPGILYVRTDIGMYRWDAANGRWIPLLDWYGNWNGYGVVIAAD 120
DB 61 GGGFVDGIVFNEGAPGILYVRTDIGMYRWDAANGRWIPLLDWYGNWNGYGVVIAAD 120
QY 121 PINTNKVAAVGMVYNSWDNDGAILRSSDQATWQITPLPFKLGGMNPGRMGERLAVD 180
DB 121 PINTNKVAAVGMVYNSWDNDGAILRSSDQATWQITPLPFKLGGMNPGRMGERLAVD 180
QY 181 PNNDNILYFGAPSGKGLWRSYVTDIGMYRWDAANGRWIPLLDWYGNWNGYGVVIAAD 240
DB 181 PNNDNILYFGAPSGKGLWRSYVTDIGMYRWDAANGRWIPLLDWYGNWNGYGVVIAAD 240
QY 241 FDKSSSSLGQASKTIFVGVADPNPNVFWRSRDCGATWQAVPGAPTGFIPHKGVFDPVNHVL 300
DB 241 FDKSSSSLGQASKTIFVGVADPNPNVFWRSRDCGATWQAVPGAPTGFIPHKGVFDPVNHVL 300
QY 301 YIATSNTPGPGYDSSGDKWKFSTVSGTWTRISPVSTDTANDYFGYSGLTIDRQHPNTIM 360
DB 301 YIATSNTPGPGYDSSGDKWKFSTVSGTWTRISPVSTDTANDYFGYSGLTIDRQHPNTIM 360
QY 361 VATQISWMPDTIIFRSTDCGATWTRIDWTSYPNRSRYVLDISAEPWLTFFGVQPNPVP 420
DB 361 VATQISWMPDTIIFRSTDCGATWTRIDWTSYPNRSRYVLDISAEPWLTFFGVQPNPVP 420
QY 421 SPKLGWDEAMAIIDPFNSDRMLYGTGATLYATNDLTWDSGGQIHIAPMVKGLEETA VND 480
DB 421 SPKLGWDEAMAIIDPFNSDRMLYGTGATLYATNDLTWDSGGQIHIAPMVKGLEETA VND 480
QY 481 LISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTGSVDYAEINPSIIVRAGSF 540
DB 481 LISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTGSVDYAEINPSIIVRAGSF 540
QY 541 DPSSQPNDRHVAFTSDGCKNWFQSGPGVTTGGTVAASADGSRFVWAPGPGQPVVYAV 600
DB 541 DPSSQPNDRHVAFTSDGCKNWFQSGPGVTTGGTVAASADGSRFVWAPGPGQPVVYAV 600
QY 601 GFNGSMAASQGVPAANAQIRSDRNPVFWRSRDCGATWQAVPGAPTGFIPHKGVFDPVNHVL 660
DB 601 GFNGSMAASQGVPAANAQIRSDRNPVFWRSRDCGATWQAVPGAPTGFIPHKGVFDPVNHVL 660
QY 661 VMFHAVPGKEGDLWLAASGLYHSTNGGSSWSAITGVSSAVNNGFCKSPGSSYPVAVV 720
DB 661 VMFHAVPGKEGDLWLAASGLYHSTNGGSSWSAITGVSSAVNNGFCKSPGSSYPVAVV 720
QY 721 GTIGGVTAAYRSDDCGTTWVINDQHOYGNWQAITGDHANLRVYIGTNGRGIVYGD 780
DB 721 GTIGGVTAAYRSDDCGTTWVINDQHOYGNWQAITGDHANLRVYIGTNGRGIVYGD 780
QY 781 GGAPSGSPSVSPSASPSSPSPPSSPSPPSSPSPPSSPSPPSSPSPPSSPSPPSSPSPP 840
DB 781 GGAPSGSPSVSPSASPSSPSPPSSPSPPSSPSPPSSPSPPSSPSPPSSPSPPSSPSPP 840

```







Db 421 APWKGLSETAVNDLISPPSGAPLISALDGLGGFTHADVTAVPSTIFTSPVFTTGSVDY 480  
Qy 527 AELNPSIIIVRAGSFPSSQPNDRHVAFTSDGGKWFQSGSEGGVTTGTTVAASADGSRFV 586  
Db 481 AELNPSIIIVRAGSFPSSQPNDRHVAFTSDGGKWFQSGSEGGVTTGTTVAASADGSRFV 540  
Qy 587 WAPGDPGPVYVAVGFGNSWAASQGVPANAOIRSDRVNPKTFYALNSGTFFYRSTDDGGVTF 646  
Db 541 WAPGDPGPVYVAVGFGNSWAASQGVPANAOIRSDRVNPKTFYALNSGTFFYRSTDDGGVTF 600  
Qy 647 QPVAAGLPSSGAVGMFHAVPKGKGLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 706  
Db 601 QPVAAGLPSSGAVGMFHAVPKGKGLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 660  
Qy 707 KSAPGSSYPVAVFVGTIGGVTGAYRSDDCGTTWVLINDDQHOYGNWGQAITGDHANLRRV 766  
Db 661 KSAPGSSYPVAVFVGTIGGVTGAYRSDDCGTTWVLINDDQHOYGNWGQAITGDHANLRRV 720  
Qy 767 YIGTNGRGIVYGDIGGAPSG 786  
Db 721 YIGTNGRGIVYGDIGGAPSG 740  
RESULT 5  
US-10-155-400-3  
; Sequence 3, Application US/10155400  
; Publication No. US20030108988A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
; FILE REFERENCE: NREL 01-36A  
; CURRENT APPLICATION NUMBER: US/10/155,400  
; CURRENT FILING DATE: 2002-10-22  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 740  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; OTHER INFORMATION: Catalytic domain GH74  
US-10-155-400-3  
Query Match 78.6%; Score 4036; DB 14; Length 740;  
Best Local Similarity 100.0%; Pred. No. 1.5e-218;  
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 47 ATTQPYTWSNVAIGGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 106  
Db 1 ATTQPYTWSNVAIGGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60  
Qy 107 NNWGYNGVVSIAADPINTNKVAAVGMVNTNSWDNDGAILRSSDQATWQITPLPFKLG 166  
Db 61 NNWGYNGVVSIAADPINTNKVAAVGMVNTNSWDNDGAILRSSDQATWQITPLPFKLG 120  
Qy 167 NMPGRGMGERLAVDPNNNDILYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDT 226  
Db 121 NMPGRGMGERLAVDPNNNDILYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDT 180  
Qy 227 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNNPVFWSRGGGATWQAVPGAPTGF 286  
Db 241 IPHKGVFDPVNHVLIATNSGTGPDGSSGDVWKFSVTSGTWTRISPVSTDTANDYFGY 300  
Qy 347 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDDGGATWTRINDWTSYPNRSRLRYVLDSAE 406

Db 301 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDDGGATWTRINDWTSYPNRSRLRYVLDSAE 360  
Qy 407 PMLTFGVQPNPVPSPKLGWMDMAIDPENSRLMLYGTGATLYATNDLTKWDSGGQIHI 466  
Db 361 PMLTFGVQPNPVPSPKLGWMDMAIDPENSRLMLYGTGATLYATNDLTKWDSGGQIHI 420  
Qy 467 APWKGLSETAVNDLISPPSGAPLISALDGLGGFTHADVTAVPSTIFTSPVFTTGSVDY 526  
Db 421 APWKGLSETAVNDLISPPSGAPLISALDGLGGFTHADVTAVPSTIFTSPVFTTGSVDY 480  
Qy 527 AELNPSIIIVRAGSFPSSQPNDRHVAFTSDGGKWFQSGSEGGVTTGTTVAASADGSRFV 586  
Db 481 AELNPSIIIVRAGSFPSSQPNDRHVAFTSDGGKWFQSGSEGGVTTGTTVAASADGSRFV 540  
Qy 587 WAPGDPGPVYVAVGFGNSWAASQGVPANAOIRSDRVNPKTFYALNSGTFFYRSTDDGGVTF 646  
Db 541 WAPGDPGPVYVAVGFGNSWAASQGVPANAOIRSDRVNPKTFYALNSGTFFYRSTDDGGVTF 600  
Qy 647 QPVAAGLPSSGAVGMFHAVPKGKGLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 706  
Db 601 QPVAAGLPSSGAVGMFHAVPKGKGLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 660  
Qy 707 KSAPGSSYPVAVFVGTIGGVTGAYRSDDCGTTWVLINDDQHOYGNWGQAITGDHANLRRV 766  
Db 661 KSAPGSSYPVAVFVGTIGGVTGAYRSDDCGTTWVLINDDQHOYGNWGQAITGDHANLRRV 720  
Qy 767 YIGTNGRGIVYGDIGGAPSG 786  
Db 721 YIGTNGRGIVYGDIGGAPSG 740  
RESULT 6  
US-10-155-400-6  
; Sequence 6, Application US/10155400  
; Publication No. US20030108988A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
; FILE REFERENCE: NREL 01-36A  
; CURRENT APPLICATION NUMBER: US/10/155,400  
; CURRENT FILING DATE: 2002-10-22  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 740  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; OTHER INFORMATION: Catalytic domain GH74  
US-10-155-400-6  
Query Match 78.6%; Score 4036; DB 14; Length 740;  
Best Local Similarity 100.0%; Pred. No. 1.5e-218;  
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 47 ATTQPYTWSNVAIGGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 106  
Db 1 ATTQPYTWSNVAIGGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60  
Qy 107 NNWGYNGVVSIAADPINTNKVAAVGMVNTNSWDNDGAILRSSDQATWQITPLPFKLG 166  
Db 61 NNWGYNGVVSIAADPINTNKVAAVGMVNTNSWDNDGAILRSSDQATWQITPLPFKLG 120  
Qy 167 NMPGRGMGERLAVDPNNNDILYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDT 226  
Db 121 NMPGRGMGERLAVDPNNNDILYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDT 180  
Qy 227 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNNPVFWSRGGGATWQAVPGAPTGF 286  
Db 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNNPVFWSRGGGATWQAVPGAPTGF 240



QY 287 IPHKGVEPDPVNHVLIATISNTGGPYDSSGDMVKFESVTSCTWTTRISPVPSDTDTANDYFGY 346  
DB 241 IPHKGVEPDPVNHVLIATISNTGGPYDSSGDMVKFESVTSCTWTTRISPVPSDTDTANDYFGY 300  
QY 347 SGLTIDRQHPNTIMVATQISWNPDTIIPRSTDDGATWTRIMDWTSYNRSLRYVLDISAE 406  
DB 301 SGLTIDRQHPNTIMVATQISWNPDTIIPRSTDDGATWTRIMDWTSYNRSLRYVLDISAE 360  
QY 407 PWTFTGQPNPVPSPKLGWMDMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 466  
DB 361 PWTFTGQPNPVPSPKLGWMDMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420  
QY 467 APWKGLBETAVNDLISPPSGAPLISALDGGTHADVTAVPSTIPTSVPFTTGTSDY 526  
DB 421 APWKGLBETAVNDLISPPSGAPLISALDGGTHADVTAVPSTIPTSVPFTTGTSDY 480  
QY 527 AELNPSIIVRAGSFDPSSQPNDRHVARSTDDGKNWFGQSEPGVTTGGTVAASADGSRFV 586  
DB 481 AELNPSIIVRAGSFDPSSQPNDRHVARSTDDGKNWFGQSEPGVTTGGTVAASADGSRFV 540  
QY 587 WAPDGPQPVVYVGVFGNSWAASQGVANAGIQRSDRVNPKTFYALSNCTFYRSTDDGVTF 646  
DB 541 WAPDGPQPVVYVGVFGNSWAASQGVANAGIQRSDRVNPKTFYALSNCTFYRSTDDGVTF 600  
QY 647 QPVAAGLPSSGAVGMFHAVPGKEGDLMLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 706  
DB 601 QPVAAGLPSSGAVGMFHAVPGKEGDLMLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 660  
QY 707 KSAPOSSYPVAVVGTIGVGTAYRSDCGTTWLLINDDQHOYGNWQAIGTDHANLRRV 766  
DB 661 KSAPOSSYPVAVVGTIGVGTAYRSDCGTTWLLINDDQHOYGNWQAIGTDHANLRRV 720  
QY 767 YIGTNGRGIVYGDIGGAPSG 786  
DB 721 YIGTNGRGIVYGDIGGAPSG 740

RESULT 7  
US-10-156-761-9395  
; Sequence 9395, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; PRIOR FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 9395  
; LENGTH: 882  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-9395  
Query Match 48.3%; Score 2478; DB 14; Length 882;  
Best Local Similarity 50.4%; Pred. No. 7e-131;  
Matches 473; Conservative 128; Mismatches 227; Indels 110; Gaps 14;  
QY 23 TASFAVAAL--GVLPATATAPAHAAATQPTWNSVAIGGGGVYDGIYFNEGACILYV 80  
DB 8 TAVLALAGLPATPPALAAAPTATTAADTYSWKNAVDDGGVFPVFNRSKKNLAYA 67

QY 81 RTDIGMYRWAANGRWIPLLDWGMNNGVNGVVSIAADPINTNKNVAAVGMVYTNWD 140  
DB 68 RTDIGMYRWAANGRWIPLLDWGMNNGVNGVVSIAADPINTNKNVAAVGMVYTNWD 127  
QY 141 NDGAILRSDGATWQITPLPFKLGNNMPGRGMRGLAVDPNNNLLYFGAPSGKGLWRS 200  
DB 128 GNGAVLRSDGASWQKTDLPKLGNNMPGRGMRGLAVDPNNRNSVLYLGAAPSGKGLWRS 187  
QY 201 TDSGATWQMTNFPDVGTYIANPTDTTGYOSDIQGVVWVAFDKSSSSLLQQAOKTIFVGV 260  
DB 188 TDSGASWQMTNFPDVGTYIANPTDTTGYOSDIQGVVWVAFDKSSSSLLQQAOKTIFVGV 247  
QY 261 DPNNPFFWSDGATWQAVPGAPTGFIPHKGVDFPNVHLYIATSNCTGPGYDSSGDMVK 320  
DB 248 DKONSIVYSTDAGATWSRLAGQPTGHLAHKGVLDAAANGCLYLAYSCKGPGYDGGKQLWR 307  
QY 321 FSVTSGTWTRISPVPSDTDTANDYFGYSGLTIDRQHPNTIMVATQISWNPDTIIFRSTDDG 380  
DB 308 YTTKTGTWNTISPVAAEDT---YGFSGLTVDROHPGTVMATAYSSWNPDTQLFRSTDSG 364  
QY 381 ATWTRIMDWTSYNRSLRYVLDISAEPMWLTFCVQPNPVPSPKLGWMDMAIDPFNSDR 440  
DB 365 GTWTKAWDYTSYPSNSRFTMDVSSPMLTGANPAPPEQTPLKGMWTSLELDPPDSAR 424  
QY 441 MLYGTGATLYATNDLTKWDSGGQIHIAPMVKGLIETAVNDLISPPS--GAPLISALDGLG 499  
DB 425 MMYGTAIVYGTNDLITNWDSSGQFTIKPMARGLEETAVNDLASPPSGGQALFSALDGLG 484  
QY 500 FTHADVAVPSTIPTSVPFTTGTSDYAEINPISIIVRAGSFDPSSQPNDRHVARSTDDGV 559  
DB 485 FRHTDLTTPVPSLMYSPNFTTSTSLDYAETDPGTVVVGNLD--SGP--HVAFTSDNGA 539  
QY 560 NWFQSEPGVTTGGTVAASADGSRFVWAPDGPQPVVYVGVFGNSWAASQGVANAGIQR 619  
DB 540 NWFAGADPSGSGGCTVAASDGRFVWSPAGTG--VQYTTGFTGTSWASAGLPAGAIVE 597  
QY 620 SDRVNPKTFFYALSNCTFYRSTDDGVTFQPVAA--GLPSSGAVGMFHAVPGKEGDLMLAAS 678  
DB 598 SDRVDPKTFYFGKSGRFVYSSDGGATFTASAATGLPSGDS--VRFKALPCKGDIWL 655  
QY 679 S-----GLYHSTNGSSWSAITGVSSAVNVGFKSAPSSYPAVFFVGTIGGVTGAYRSD 733  
DB 656 ASDGAYGLWHSTDDGGAFTKLATVDQADTIGFKAATGASYQTLTYSKIGGVRGIFRST 715  
QY 734 DCGTTWLLINDDQHOYGNWQAIGTDHANLRRVYIGTNGRGIVYGDIGGAPSGSPSPSV 793  
DB 716 DKGASWTRVNDDAHQWGTGAATIGDPRVYGVYVSTNGRGIVYGDITAGSSDGG----- 769  
QY 794 PSASPSLSPPSP 853  
DB 770 ----- 769  
QY 854 SSPSSSP 913  
DB 770 ---GTEPAT-----GACTVYIRITNWSGGFQ--ADVQLANTGSTANDGWSLG-- 814  
QY 914 FTRDGSSTLVNCDMAAIGCG-----NIRASFGS 943  
DB 815 SFGDGGQEVTLNWSVAQAAGSGVTAANLAWNGRVAAGS 852

RESULT 8  
US-09-917-376-7  
; Sequence 7, Application US/09917376  
; Publication No. US20040038334A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOGLYTICUS  
; FILE REFERENCE: 40197.40501







```
RESULT 10
US-10-420-191-2
; Sequence 2, Application US/10420191
; Publication No. US20040067569A1
; GENERAL INFORMATION:
; APPLICANT: Rev. Michael W.
; APPLICANT: Zaretsky, Elizabeth J.
; APPLICANT: Haas, Jeffrey A.
; TITLE OF INVENTION: Polypeptides Having Xyloglucanase Activity And Nucleic Acids
; TITLE OF INVENTION: Encoding Same
; FILE REFERENCE: 10210.200-US
; CURRENT APPLICATION NUMBER: US/10/420,191
; CURRENT FILING DATE: 2003-04-18
; PRIOR APPLICATION NUMBER: US 60/373,987
; PRIOR FILING DATE: 2002-04-19
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-10-420-191-2

Query Match 31.7%; Score 1625.5; DB 15; Length 838;
Best Local Similarity 41.7%; Pred. No. 5.5e-83;
Matches 350; Conservative 131; Mismatches 296; Indels 63; Gaps 23;

QY 25 SPFAVAALGLPIALTAGPAHAATQPYTWSNVAI-GGGGFGDGIIVFNEGAPGILYVRTD 83
DB 4 SRVLALVGV-----IPAHAA-----PSWKNVKLGCGGFGVPGIIFHPKTKGVAYARTD 53

QY 84 IGMWRWDAANGRWIPLLDWV-----GMNNGYGVVSIADPINTNKVAAVGMVNTNSWD 139
DB 54 IGLYLRLN-ADDSWTAVTDTGIADNAGWNW---GIDAVALDQDDQKVAAVGMVNTNSWD 109

QY 140 PNDGAILRSSDQATWQITPLPKLGGNMPGMRGLAVIDPNDNLIYFGAPSGKGLWR 199
DB 110 PSNGAIIRSSDRGATWSFTNLFPKVGNNPGRGAGERLAVIDPANSNIIYFGARSGNGLWK 169

QY 200 STDSGATWSOMTFPDVGTIYIANPTDTTGYQSDIQGVVWVAIPDKSSSSLGQASKTIFVGV 259
DB 170 STDGGTFSKSSFTATGIYIPDSDSGYNSDKQLMWVTFDSTSTTGATSRIFVGT 229

QY 260 ADP-NNPVFWSRDGGATQAVPGAPTGFIPHRGVFDPVNVHLYIATSNTPGYPDSSGDV 318
DB 230 ADNITASVYVSTNAGSTWSAVPGQPKYFPHKAKLQPAEKALYLTYSDCGTGYPDGLGV 289

QY 319 WFSVTSQGTWTRISVPSTDTANDYFGYSGLIIDROHPNTIIVATQISWPDITIPRSTD 378
DB 290 WRYDIAGGTWKDITFVSGSDL---YFGGGLGLDLQKPGTLVVASLNSWPPDAQLPRSTD 346

QY 379 GGNATRWDTSYNRSRLRYVLDISAEPLTFG-VQPNPPVPS---PKLGWMDAMAI 433
DB 347 SGTWSPIMAWASYPETETYYISITPKAPWIKNFIDVTSESPDGLIIRKLGWMIESLEI 406

QY 434 DPNSDRMLYGTGATLYATNDITKWDSCGQIHIAPMVKGLIETAVNDLISPPSGAPLISA 493
DB 407 DPTDSNHWLYGTGWTIFGCHDLTNWDTNRNVSITQSLADGIEFVSQDLASAPGSELLAA 466

QY 494 LGDLGFTTHA---DVTAVPSTIFTSPVFTTGTSDVVAELNPSIIVRAGSDFDSSQPNDRH 550
DB 467 VGDNGFTFASRNDLGTSPQVWATPTWATSTSDVYAGNSVKSVVRVGNVGTAGTQQ----- 521

QY 551 VAFSTGGKWNFGQSGEPGVTGGTVAASADSRFWAFCDQPPQVYVAVGFGNSWAASQ 610
DB 522 VAISDGGATWSIDYAADTSMNGGTVAISADGDTILWSTASSG---VQRSQFGSGFASVS 578

QY 611 GVPANAQIRSDRVNPKTFYALNGTFRYRTDGGVTFPQVAAAGLPSGSGVGM--FHAVPG 668
DB 579 SLPAGAVIASDKTNSVVFAGSGSTFVYSKDTGSSF---TRG-PKLGSGATIRDIAAHT 634

QY 669 KEGDLMAASSGLYHSTNGGSSWSAI-TGVSSAVNVVFGKSGAPSSYPAVFVVTGGVTT 727

Db 635 TAGTLYVSTDVGIIFRSTDSGTTFGQVSTALTNTTYQIALGVGS-GSNW-NLYAFET--GPS 690
QY 728 GA--YRSDDCGTTWVLINDDQHQYGNWGOAITGDHANLRRVYIGTNGRGIVY--GDICGA 783
Db 691 GARLYASGDSGASWTDIQSQGFGSIDSTKVAGSGTAGGVYGTNGRGVYAGTGVGG 750
QY 784 PSGSPSPVSPASPSLSPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 843
Db 751 TGGT-----SSSTKQSSSTSSASSSTTLRSSVWSTTRASTVTSRTSSAAGTGS 801

RESULT 11
US-10-026-994-2
; Sequence 2, Application US/10026994
; Publication No. US20030113732A1
; GENERAL INFORMATION:
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Goedegebuur, Frits
; APPLICANT: Ward, Michael
; APPLICANT: Yao, Jian
; TITLE OF INVENTION: ECVI Endoglucanase and Nucleic Acids
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: GC698
; CURRENT APPLICATION NUMBER: US/10/026,994
; CURRENT FILING DATE: 2002-04-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 818
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-10-026-994-2

Query Match 31.2%; Score 1604; DB 14; Length 818;
Best Local Similarity 41.8%; Pred. No. 8.6e-82;
Matches 340; Conservative 128; Mismatches 291; Indels 54; Gaps 21;

QY 52 YTWSNVAI-GGGGFGDGIIVFNEGAPGILYVRTDTIGMYRWDAANGRWIPLLDWV---GW 106
DB 2 FSWKNVKGCGGFGVPGIIFHPKTKGVAYARTDIGGLYRLN-ADDSWTAVTDTGIADNAGW 60

QY 107 NMNGYGVVSIADPINTNKVAAVGMVNTNSWPNDCGAILRSSDQATWQITPLPKLGG 166
DB 61 HNW---GIDAVALDQDDQKVAAVGMVNTNSWPNDCGAILRSSDQATWQITPLPKLGG 117

QY 167 NMEGRGMRGLAVIDPNDNLIYFGAPSGKGLWRSTDSGATWSOMTNPPDVGTIYIANPTDT 226
DB 118 NMEGRGAGERLAVIDPANSNIIYFGARSGNGLWKSTDCGVTFSKVSSFTATGTIYIPDPSDS 177

QY 227 TGYQSDIQGVVWVAIPDKSSSSLGQASKTIFVGVADP-NNPVFWSRDGGATQAVPGAPTG 285
DB 178 NGYNSDKQGLMWVTFDSTSTTGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQPKG 237

QY 286 FIPKGVFPVNVHLYIATSNTPGYPDSSGDVWKFVSQVTSQWTRISVPSTDTANDYFG 345
DB 238 YFPHKAKLQPAEKALYLTYSDCGTGYPDGLGVVWRYDIAGGTWKDITFVSGSDL---YFG 294

QY 346 YSGLTIDROHPNTIIVATQISWPDITIFRSTDGGATWTRIDWTSYPNRSLRYVLDISA 405
DB 295 FGGGLGLDLQKPGTLVVASLNSWPPDAQLFRSTDSGTTSPSPIMAWASYPETETYYISITPK 354

QY 406 EPLMTFG-VQPNPPVPS-----PKLGWMDAMAIIDPFNSDRMLYGTGATLYATNDITKWD 460
DB 355 APWIKNFIDVTSESPDGLIIRKLGWMIESLEIDPTDSNHWLYGTGWTIFGCHDLTNWDT 414

QY 461 GGGIHIAPMVKGLIETAVNDLISPPSGAPLISALGDLGGFTHA---DVTAVPSTIFTSPV 517
DB 415 RHNVSITQSLADGIEFVSQDLASAPGSELLAAVGGDNGTFTFASRNDLGTSPQVWATPT 474

QY 518 FTTGTSVDVVAELNPSIIVRAGSDFDSSQPNDRHVAFTSDGKWNFGQSGEPGVTGGTVA 577
DB 475 WATSTSDVYAGNSVKSVVRVGN-----TAGTQVAISSDGGATWSIDYAADTSMNGTVA 528
```



QY 578 ASADGSRFVWAPGDPGPVYVAVFGNSWAASQGVPAQAQIRSDRVNPKTFYALNSNGTFY 637  
 Db 529 YSADGDTILNSTASSG---VQRSPQGSFASVSLSLPAQAVIASDKTNSFYAGSGSTFY 585  
 QY 638 RSTDGGVTFQPVAAAGLSPSSGAVGM--FHAVPGKEGDLWLAASSGLYHSTNGGSSWSAI- 694  
 Db 586 VSKDTGSGSF--TRG-PKLSAGTIRIDIAAHTTTAGTLYVSTDVGIIFRSTDSGTTFGQVS 641  
 QY 695 TGVSAAVNVGFGKAPGSSPAPVAVGTIGVTCG--YRSDDCGTTWVLINDDQHOYGNW 752  
 Db 642 TALNTNTYQIALGVGS-GSNW-NLYAFGT--GPSGARLYASGDSGASWTDIQSGQGFSGID 697  
 QY 753 GOALTGPHANLRRVYVIGNNGRGIVY--GDIGGAPSGSPSPSPVSPSASPSLSPSPSPSSP 810  
 Db 698 STKVAGSGTAGQVYVGTNGRGVYAGTVGGGTGTT-----SSSTKQSSSTSSA 748  
 QY 811 SPSPSPSSPSSPSPSPSPSPSPSPSPSPSPSAS 843  
 Db 749 SSSTTLRSSVVSTTRASTVTSRTSSAAGPTGS 781

RESULT 12

US-10-156-761-10111  
 ; Sequence 10111, Application US/10156761  
 ; Publication No. US20030119018A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OMURA, SATOSHI  
 ; APPLICANT: IKEDA, HARUO  
 ; APPLICANT: ISHIKAWA, JUN  
 ; APPLICANT: HORIKAWA, HIROSHI  
 ; APPLICANT: SHIBA, TADAYOSHI  
 ; APPLICANT: SAKAKI, YOSHIYUKI  
 ; APPLICANT: HATTORI, MASAHIRA  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-262  
 ; CURRENT APPLICATION NUMBER: US/10/156,761  
 ; CURRENT FILING DATE: 2002-05-29  
 ; PRIOR APPLICATION NUMBER: JP 2001-204089  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; PRIOR FILING DATE: 2001-08-02  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 10111  
 ; LENGTH: 739  
 ; TYPE: PRT  
 ; ORGANISM: Streptomyces avermitilis  
 US-10-156-761-10111

Query Match 28.1%; Score 1442; DB 14; Length 739;  
 Best Local Similarity 40.8%; Pred. No. 9.7e-73;  
 Matches 328; Conservative 107; Mismatches 272; Indels 96; Gaps 25;  
 QY 7 IRLTMRRLVSLAATASFAVAALGVLPATITASPAAHAATT---QPYTWSNVAIGGG 63  
 Db 1 MTRPRPRTV--LAGTAA---AALATVP--AVGQAAHAETTAGPSYRWRNAVIGTG 53  
 QY 64 FVDGIVFNEGAPGILYVTRDTIGMYRMDAANGRIPLDWMYGNWYGVVSTAADPIN 123  
 Db 54 FVTGVLFHPSVRGLAYARTDIGGAYRWDGRTPLDHLGWDNLLGVEAMAVDPHT 113  
 QY 124 TNKWAAGVMTNSWDPNDGAILRSSDQATWQITPLPFKLGNNPGRGMRERLAVDPNN 183  
 Db 114 PDRLYLAVGTYAQSWAGN-GAVLRSEDRTGATWTRTDLTVKLGGNEDRGAGERLLVDP 172  
 QY 184 DNILYFGAPSKGLWRSTDGATSQMTNFFDVGTYIANPTDTTGYQSDIQGVVWVAFDK 243  
 Db 173 SDTLWLGT-RHDLGKSTDRGATWAAATAFP-----AKANSSGQGVF----- 214  
 QY 244 SSSSLGQASKTIFVGVADPNPNP-----VFWSRDGGATWQAVPGAPTGF---IPHKGVDP 295  
 Db 215 ----LVAAGRTVYAGWGDGDTSGTANLYRTAD-GTTWGAVGRPSGTSAKVPLRAAYDT 269

QY 296 VNHVLYIATSNLTGGPYDSSGSDVWKFVSTSGTWTTRISPV-----PSTDTANDVFGYSGLTI 351  
 Db 270 HTRELYVYTGAPGPGGQSDGSHKLRATATGTWTEVTPVKPGTTSDDGADTFAYGGVAV 329  
 QY 352 DRQPNNTIMVATQISWPDITIFRSTDCGATWTRINDWTSYNRSRYVLDISAEPWLTF 411  
 Db 330 DARRPGTLVSTNNRWADGTVFRSTDCGRTWTSKLD-----AAVFDVSETPFLDW 380  
 QY 412 GVQPNPPVPSPKLGWMDMAIDPNSDRMLYGTGATLYATNDLTKWDSGGQIHIAPMVK 471  
 Db 391 GDD-----KPKFGWIIQALAVDPDSQHVYGTGATLYGTRDLKRW-----APRIR 426  
 QY 472 GLEETAVNDLISPPSG-APLISALGDLGGFTHADVTAVEST-IFTSVPVTTGTSVDYAE 529  
 Db 427 GLEESAVRQLISPPVGEAHLISGLDIGVMYHERLTASPSRGWATNPVFGSATGLAQAAA 486  
 QY 530 NPSIIVRAGSPDPSSQPNDRHVAFTSDGKNW--PQSEPPGGVTTGCTTAAASADGSRFVW 587  
 Db 487 RPAYVVRTGWDGHNG-----AYSHDGGRTWAPFEAQPIADKAPGIATSDAGGTLW 540  
 QY 588 A----PGDPGQPVVYA---VGFNGSWAASQGVPAQAQIRSDRVNPKTFYA--LSNGTFY 637  
 Db 541 SFVHWG-----TTYAHRSTNDGASWSEVSPFKGATPVADPADTRFYAYDFDNGTLY 595  
 QY 638 RSTDGGVTFQPVAAAGLSPSSGAVGMFHAVPGKEGDLWLAAS-SGLYHSTNGGSSWSAITG 696  
 Db 596 ASTDSGRSFTARAGCLP-SGDSQFKLVAAPGRSGDLWLSAKWNGLYRSTDDGDTFARIDS 654  
 QY 697 VSSAVNVGFGKAPGSSYPAVVFVGTIGVTCAYRSDDCGTTWVLINDDQHOYGNWQAI 756  
 Db 655 CWASYTLGLGKAADGADYPAIYQVGSSTETITAYRSDDAARTWVRINDDAHOMGWIGEAV 714  
 QY 757 TGDHANLRRVYVGTNGRGIVYGD 779  
 Db 715 VGDPRHGRVYLATNGRGIQYGE 737

RESULT 13

US-10-395-241-12  
 ; Sequence 12, Application US/10395241  
 ; Publication No. US20040038367A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: YAOI, Katsuro  
 ; APPLICANT: MITSUISHI, Yasushi  
 ; TITLE OF INVENTION: NOVEL XYLOGLYCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE  
 ; FILE REFERENCE: Q73756  
 ; CURRENT APPLICATION NUMBER: US/10/395,241  
 ; CURRENT FILING DATE: 2003-03-25  
 ; PRIOR APPLICATION NUMBER: JP 2002-83433  
 ; PRIOR FILING DATE: 2002-03-25  
 ; NUMBER OF SEQ ID NOS: 19  
 ; SOFTWARE: Patent in version 3.2  
 ; SEQ ID NO 12  
 ; LENGTH: 812  
 ; TYPE: PRT  
 ; ORGANISM: Geotrichum sp. M128  
 US-10-395-241-12

Query Match 21.9%; Score 1125; DB 15; Length 812;  
 Best Local Similarity 34.5%; Pred. No. 7.2e-55;  
 Matches 285; Conservative 129; Mismatches 291; Indels 120; Gaps 31;  
 QY 31 ALGVLPITAITASPAHAATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVTRDIGMYRW 90  
 Db 6 SLGKALTALSILASLAVAKEHYEFKNVAIGGGYITGIVAHPKTKDLYARTDIGGAYRW 65  
 QY 91 DAANGRWPLLDWGVNNWNGVVSIAADPINTNKVWAAVGMYT-NSWDPNDGAILRSS 149  
 Db 66 DAGTSKWTPLNDFIEAQDMNINMGTSIALDPNPDRLYLAQGRYVGDEW----AAFVSE 121  
 QY 150 DQATWQITPLPFKLGNNPGRGMRERLAVDPNDNDILYFGAPSGKGLWRSTDSGATWSQ 209



Db 122 DRGQSFTIYESPPFPMGANDMGRNGERLAVNPNFNSNEVMGTRT-EGIKWSDRAKTWTN 180  
 Qy 210 MTFNPDVCTYIANPTDTTGYQSDTQGVVWVAFDKSSSLGQASKTIIFGVADPNPNPVFWS 269  
 Db 181 VTSIPDAF-----TNGIGYTS-----VIFDP-----ERNGTIYASATAPOG-MYVT 220  
 Qy 270 RDGGATQAVPGAPTGT-----PHKGVFDPVNVHLYIATSNITGPGY 311  
 Db 221 HDGVSWEFVAGQSSSWLNRTTGAPPKPSIAPOPMKVALT-----NPLYVYIADYPPGW 278  
 Qy 312 DGSSGDVWKFVSTGTWTRI-----SPVSTDTANDYFYSGLTIDROHPNTIMVATQ 364  
 Db 279 GVTFCVWRQRTSGAMDITPRVGNSSPAPVNNQTPPAGFCGLSVDTATPNRLVWIT- 337  
 Qy 365 ISWMPDTI-----IFRSTDCGATWTRIDWTS-----YPNRSLRYVLDISAEWLPFG 412  
 Db 338 LDRDPGALDSIYLSIDAGATWKTQVLSLSPNSLEGNWGHPTNAARY-KDGTVPVWLDPN 396  
 Qy 413 VQP-----NPPVPS-----KLGMMDAMAI DPNFSDRLMLYGTATLYATNDLT-----WDS 460  
 Db 397 NGPQGGYGAPHGTPGLTKFGWMSAVLIDPFNPEHLMYGTATWATDLSRVEKDW-- 454  
 Qy 461 GQOIHIAP-----MVKGLEBETAVNDLISPPSGAPLISALDGLGFTHADVTAVPSTFTSP 516  
 Db 455 -----AESWYLIQIDGIEENALSLRSPKSGAALLSGIGDISGMKHDLLTK-PQRMFGAP 507  
 Qy 517 VFTTGTSDVYAEINPSIIVRAGSPDPQNDRHVAFSTDDGKNW---FQSGEPGGVTT--- 572  
 Db 508 QPSNLDSDAAGNFNPNVVRAGSGHEYDSACARGAYATDGDGDAWTFPTCPGKNASHY 567  
 Qy 573 -GGTVAASADGRFRWAP--GDPGPVYVAVFGNSWAAASQGVPA-----NAQIRSDRV 623  
 Db 568 QGSTIADVAGSQIYVWSTKLDEQASGPMYSHDYGKTS-----VPAGDLKAQNTANVLSDKV 623  
 Qy 624 NPKTYALSNGTFRSTDTGGVTFQVAAAGLPSGAGVGMFHAVPGKEGDLWL-AASSGLY 682  
 Db 624 QDGTFTYADGKFFVSTDDGKSYAAKAGGLVT--GTSMLPVPVWAGDVWVPEGGGLF 681  
 Qy 683 HSTNGSSWSAI-TGVSSAVNVFGKS-----APGSSYPVAVFVGT--IGGVTGAYRSDDC 735  
 Db 682 HSTDFGAFTRVGTANATLVSVGAPKSDGKASAPSAVFITWGTDPKGSIDGLYRSDN 741  
 Qy 736 GTTWVLIINDQHOYGNWQQAITGDHANLRVYIGTNGRGIVYDI 780  
 Db 742 GSTWTRVNDQEHNSG-PTMIBADPKYGRVYLTNGRGIVYADL 785

RESULT 14  
 US-10-395-241-14  
 ; Sequence 14, Application US/10395241  
 ; Publication No. US20040038367A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: YAOI, Katsuro  
 ; APPLICANT: MITSUISHI, Yasushi  
 ; TITLE OF INVENTION: NOVEL XYLOGLYCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE  
 ; TITLE OF INVENTION: ENCODING THE SAME, AND METHOD OF PREPARING THE ENZYME  
 ; FILE REFERENCE: Q73756  
 ; CURRENT APPLICATION NUMBER: US/10/395,241  
 ; CURRENT FILING DATE: 2003-03-25  
 ; PRIOR APPLICATION NUMBER: JP 2002-83433  
 ; PRIOR FILING DATE: 2002-03-25  
 ; NUMBER OF SEQ ID NOS: 19  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 14  
 ; LENGTH: 789  
 ; TYPE: PRT  
 ; ORGANISM: Geotrichum sp. M128  
 US-10-395-241-14

Query Match 21.8%; Score 1117; DB 15; Length 789;  
 Best Local Similarity 35.0%; Pred. No. 28-54; Indels 120; Gaps 31;  
 Matches 281; Conservative 124; Mismatches 279;

Qy 52 YTMSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAANGRWIPLLDWVGNWGY 111  
 Db 4 YEFKVAIGGGYITGIVAHPTKOLLYARTDIGGAYRWDAGTSKWIPLDFTFQAQDNI 63  
 Qy 112 NGVSIADPINTNKVAAVGMYT-NSWDPNDGAILRSSDOGATWQITPLPKLGGNMPG 170  
 Db 64 MGTESIALDPNPNDRLLYLAQGRYVGDEW-----AFVVSSEDRGQSFTIYESPPPMGANDMG 119  
 Qy 171 RGMGERLAVDPNNDILYFAGPSGKGLMRSTDSGATWSQMTNFPDVGTVIANPTDTTGYQ 230  
 Db 120 RNNGERLAVNPNFNSNEVMGTRT-EGIKWSBRKATWNTVTSIPDAF-----TNGIGYT 172  
 Qy 231 SDIQGVWVAFDKSSSLGQASKTIIFGVADPNPNFVNSRDGATWQAVPGAPTGTI--- 287  
 Db 173 S-----VIFDP-----ERNGTIYASATAPOG-MYVTHDGGVSEVPVAGQSSWLNRT 218  
 Qy 288 -----PHKGVFDPVNVHLYIATSNITGPGYDGSDDWVWKSFSVSGTWTRI- 331  
 Db 219 TGAPFDKPSIAPOPMKVALT-----NPLYVYIADYPPGWVTFGEVWRQRTSGAWDDIT 276  
 Qy 332 -----SPVSTDTANDYFYSGLTIDROHPNTIMVATQISWMPDTI-----IFRSTDCGAT 382  
 Db 277 PRVGNSSPAPYNNQTPPAGFCGLSVDTATPNRLVWIT-LDRDPGALDSIYLSIDAGAT 335  
 Qy 383 WTRIWDWTS-----YPNRSLRYVLDISAEWLPITFGVQP-----NPPVPS-----KLG 425  
 Db 336 WKDVTQLSSPNSLEGNWGHPTNAARY-KDGTVPVWLDNFNPGQGGYGAPHGTPGLTKFG 394  
 Qy 426 WMDEAMAI DPNFSDRLMLYGTATLYATNDLT-----WDSGGQIHIAP-----MVKGLEETA 477  
 Db 395 WMSAVLIDPFNPEHLMYGTATWATDLSRVEKDW-----APSWYLIQIDGIEENA 446  
 Qy 478 VNDLISPPSGAPLISALDGLGFTHADVTAVPSTFTSPVFTTGTSDVYAEINPSIIVRA 537  
 Db 447 ILSLRSPKSGAALLSGIGDISGMKHDLLTK-PQRMFGAPQFNLSDIDAAGNFNPNVVR 505  
 Qy 538 GSFDPSSQPNDRHVAFSTDDGKNW---FQSGEPGGVTT---GGTVAASADGRFRWAP--G 590  
 Db 506 GSGGHEYDSACARGAYATDGDGDAWTFPTCPGKNASHYQGSTIADVAGSQIYVWSTKLD 565  
 Qy 591 DPGQPVYVAVFGNSWAAASQGVPA-----NAQIRSDRVNPKTFYALSNGTFRSTDTGGV 644  
 Db 566 EQASGPMYSHDYGKTS-----VPAGDLKAQNTANVLSDKVDGTFYATDGGKFFVSTDDGK 621  
 Qy 645 TFPQVAAAGLPSGAGVGMFHAVPGKEGDLWL-AASSGLYHSTNGSSWSAI-TGVSSAVN 702  
 Db 622 SYAAKAGGLVT--GTSMLPVPVWAGDVWVPEGGGLFHSITDFGASFTRVGTANATLV 679  
 Qy 703 VGFGKS-----APGSSYPVAVFVGT--IGGVTGAYRSDDCGTTWVLIINDQHOYGNWQAI 756  
 Db 680 VGAPKSDGKASAPSAVFITWGTDPKGSIDGLYRSDNGSTWTRVNDQEHNSG-PTMI 738  
 Qy 757 TGDHANLRVYIGTNGRGIVYDI 780  
 Db 739 EADPKYGRVYLTNGRGIVYADL 762

RESULT 15  
 US-10-395-241-18  
 ; Sequence 18, Application US/10395241  
 ; Publication No. US20040038367A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: YAOI, Katsuro  
 ; APPLICANT: MITSUISHI, Yasushi  
 ; TITLE OF INVENTION: NOVEL XYLOGLYCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE  
 ; TITLE OF INVENTION: ENCODING THE SAME, AND METHOD OF PREPARING THE ENZYME  
 ; FILE REFERENCE: Q73756  
 ; CURRENT APPLICATION NUMBER: US/10/395,241  
 ; CURRENT FILING DATE: 2003-03-25  
 ; PRIOR APPLICATION NUMBER: JP 2002-83433  
 ; PRIOR FILING DATE: 2002-03-25  
 ; NUMBER OF SEQ ID NOS: 19  
 ; SOFTWARE: PatentIn version 3.2



```
; SEQ ID NO 18
; LENGTH: 826
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Xyloglucan Oligosaccharide-Degrading Enzyme with Histidine Tag
US-10-395-241-18

Query Match      21.7%; Score 1114; DB 15; Length 826;
Best Local Similarity 35.0%; Pred. No. 3e-54;
Matches 281; Conservative 123; Mismatches 280; Indels 120; Gaps 31;

Qy 52 YTSNVAIGGGGFDGIVFNEGAGIILYVRTDIGMYRWDAANGRWIPLLDWVGWNNWGY 111
Db : ||| ||| ||| : ||| : ||| ||| ||| ||| ||| : ||| : ||| :
Qy 5 YEFKVAIGGGYITGIVHPKTKDLYARTDIGAYRWDAAGTSKWIFLNDPIEAQDMNI 64
Db : ||| ||| ||| : ||| : ||| ||| ||| ||| ||| : ||| : ||| :
Qy 112 NGVSIAADPINTKNVAAVGMT-NSWDPNNDGAILRSSDQAGTWQITPLPKLGGNMPG 170
Db : ||| ||| ||| : ||| : ||| ||| ||| ||| ||| : ||| : ||| :
Qy 65 MGTESIALDPNPNDRLYLAQRYVGDEW---AAFVYSEDRGQFTIYESPFPMGANDMG 120
Db : ||| ||| ||| : ||| : ||| ||| ||| ||| ||| : ||| : ||| :
Qy 171 RGMGERLAVDPNNNNILYFGAPSKGLWRSTDSGATWSQMTNFPDVGTYIANPTDTGYQ 230
Db : ||| ||| ||| : ||| : ||| ||| ||| ||| ||| : ||| : ||| :
Qy 121 RNNGERLAVNPFNEVNMGTGT-EGIKSSDRAKTWTNVTSPDAF-----TNGIGYT 173
Db : ||| ||| ||| : ||| : ||| ||| ||| ||| ||| : ||| : ||| :
Qy 231 SDIQGVVVAFPDKSSSLGQASKTIFVGVADPNPNPFWMSRDGGATWQAVPGAPTGFI--- 287
Db : ||| ||| ||| : ||| : ||| ||| ||| ||| ||| : ||| : ||| :
Qy 174 S-----VIFDP-----ERNGTIYASATAPQG-MVYTHDGGVSWEPVAGQPSWLNRT 219
Db : ||| ||| ||| : ||| : ||| ||| ||| ||| ||| : ||| : ||| :
Qy 288 -----PHKGVFDPVNVHLIYIATSNTPGPGYDGGSDVWKFVSVTSGTWTRI- 331
Db : ||| ||| ||| : ||| : ||| ||| ||| ||| ||| : ||| : ||| :
Qy 220 TGAFPPKKPASIAPOPMPKVALTP-NFLYVYADYVPGWGVTFGKVMRQNRNRTSGAWDDIT 277
Db : ||| ||| ||| : ||| : ||| ||| ||| ||| ||| : ||| : ||| :
Qy 332 -----SPVSTDTANDYFGSGLTIDROHNTIMVATQISWPPDTI----IFRSTDDGAT 382
Db : ||| ||| ||| : ||| : ||| ||| ||| ||| ||| : ||| : ||| :
Qy 278 PRVGNSSPAPYNNQTFPAGPGCLGSVDATNENRLNVLIT-LDRDPGALDSIYLTSDAGAT 336
Db : ||| ||| ||| : ||| : ||| ||| ||| ||| ||| : ||| : ||| :
Qy 383 WTRTWDWTS-----YPNRSRLYVLDISAEPLWTFGVOP-----NPPVPSP---KLK 425
Db : ||| ||| ||| : ||| : ||| ||| ||| ||| ||| : ||| : ||| :
Qy 337 WKDVTQLSSPNLGNWGHPTNAARY-KDGTVPVPLDNNPQWGGYGAPHTPGLTKFG 395
Db : ||| ||| ||| : ||| : ||| ||| ||| ||| ||| : ||| : ||| :
Qy 426 WMDRAMAIDPNSDRMLYGTGATLYATNDLTK----WDSGGQIHIAP-----MVKGLEETA 477
Db : ||| ||| ||| : ||| : ||| ||| ||| ||| ||| : ||| : ||| :
Qy 396 WMSAVLIDPNEPHELMYGTGATITWATDLSRVEKDW-----APSWYLQIDGIEENA 447
Db : ||| ||| ||| : ||| : ||| ||| ||| ||| ||| : ||| : ||| :
Qy 478 VNDLISPPSGAPLISALDGLGFTHADVTAVPSTIFTSPTTGTSTVDYAEINPSIIVRA 537
Db : ||| ||| ||| : ||| : ||| ||| ||| ||| ||| : ||| : ||| :
Qy 448 ILSLRSPKSGAALLSGIGDISCMKHDDLTK-PQRMFGAPQFSNLDSDAAGNFPNVVRA 506
Db : ||| ||| ||| : ||| : ||| ||| ||| ||| ||| : ||| : ||| :
Qy 538 GSPDPSPQPNDRHVAFSTDGCKW--FOGSEPPGVTT---GGTVAASADGSRFVWAP--G 590
Db : ||| ||| ||| : ||| : ||| ||| ||| ||| ||| : ||| : ||| :
Qy 507 GSSGHEYDSACARGAYATDGDAMTIPPTCPGMNASHYQGSTIADVASGSIQVWSTKLD 566
Db : ||| ||| ||| : ||| : ||| ||| ||| ||| ||| : ||| : ||| :
Qy 591 DPGQPVVYAVGVFGNSWAASQGVPA-----NAQIRSDRVNPKTFYALSNGTFFYSTDGGV 644
Db : ||| ||| ||| : ||| : ||| ||| ||| ||| ||| : ||| : ||| :
Qy 567 EQASGPWYSHDYGKTWS-----VPAGDLKAQANVLSDKVQDCTFYATDGGKFFVSTDGK 622
Db : ||| ||| ||| : ||| : ||| ||| ||| ||| ||| : ||| : ||| :
Qy 645 TFQPVAAGLPSSGAVGVNMFHVPKKEGLWL-AASSGLYHSTNGSSWSAI-TGVSSAVN 702
Db : ||| ||| ||| : ||| : ||| ||| ||| ||| ||| : ||| : ||| :
Qy 623 SYAAKAGLAVT--GTSLMPAVNPVWAGDVWVPVEGGLFHSSTDFGASFTRVGTANATLVS 680
Db : ||| ||| ||| : ||| : ||| ||| ||| ||| ||| : ||| : ||| :
Qy 703 VGFGKS-----APGSSYPAVFVVT--IGGVYGYRSDDCGTTWVLINDDQHOYGNWQAI 756
Db : ||| ||| ||| : ||| : ||| ||| ||| ||| ||| : ||| : ||| :
Qy 681 VGAPKSKSDGKKASAPSAVFITWTDKPGSDIGLYRSDDNGSTWTRVNDQEHNYSG-PTMI 739
Db : ||| ||| ||| : ||| : ||| ||| ||| ||| ||| : ||| : ||| :
Qy 757 TGDHANLRVYVYGTNGRGIVYVDI 780
Db : ||| ||| ||| : ||| : ||| ||| ||| ||| ||| : ||| : ||| :
Qy 740 EADPKYGRVYLTNGRGIVYADL 763
Db : ||| ||| ||| : ||| : ||| ||| ||| ||| ||| : ||| : ||| :
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 5, 2005, 07:09:01 ; Search time 12598.3 Seconds  
(without alignments)  
3680.789 Million cell updates/sec

Title: US-09-917-376-1

Perfect score: 5135

Sequence: 1 MDRSENRLTWRSSRLVSL.....RASFGSVNPATPTADTYLQX 957

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2.1/USPTO.spool/US09917376/runat\_04102005\_164328\_26897/app.query.fasta\_1.2524  
-DB=GenEmbl -OFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09917376 @CGN 1 1 15287 @runat\_04102005\_164328\_26897 -NCPU=6 -ICPU=3  
-NO WMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOB=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*

1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sv.\*  
13: gb.un.\*  
14: gb\_vl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5134	100.0	2869	6	AX700058 Sequence
2	2493	48.5	296500	1	SC0939128
3	2487	48.4	299800	1	AP005028 Streptomy
4	2419.5	47.1	3262	1	AF078038s1

5	2174	42.3	2823	6	AX565635 Sequence
6	2037	39.7	12732	1	AE007608 Clostridi
7	2007.5	39.1	2950	1	AJ585344 Clostridi
8	1784	34.7	2579	8	AJ292929 Agaricus
9	1782.5	34.7	2745	8	AB015511 Aspergill
10	1690	32.9	3959	8	AY040839 Aspergill
11	1633	31.8	2724	8	AK110506 Oryza sat
12	1631.5	31.8	2517	8	AY281371 Hypocrea
13	1570	30.6	299550	1	AP005031 Streptomy
14	1179.5	23.0	14520	1	AE011809 Xanthomon
15	1170.5	22.8	11910	1	AE012276 Xanthomon
16	1161	22.6	13278	1	AE001712 Thermotog
17	1125.5	21.9	2646	6	AX924429 Sequence
18	1125.5	21.9	2716	8	AB089343 Geotrichu
19	1124	21.9	2481	6	AX924435 Sequence
20	1117	21.8	2367	6	AX924431 Sequence
21	993.5	19.3	2755	8	AB116528 Geotrichu
22	823	16.0	3687	8	AX700036 Sequence
23	702	13.7	2289	6	AX700050 Sequence
24	460.5	9.0	590	8	AB1534359 Agaricus
25	431.5	8.4	5437	1	CASR69XYN2
26	429	8.4	3365	6	AX700025 Sequence
27	409	8.0	11707	1	AF078737 Caldicell
28	409	8.0	11707	6	E35099 Truncated c
29	407	7.9	299175	1	AP005023 Streptomy
30	402.5	7.8	3237	1	AF200304 Caldicell
31	398.5	7.8	5513	1	AT286105 A.thermophi
32	397.5	7.7	5439	1	CDCELA Caldcoellum
33	391.5	7.6	5284	1	CSU16308 Caldcoellum
34	389	7.6	24208	1	AY281357S2 Unculture
35	388.5	7.6	202301	1	AE017286 Desulfocov
36	387	7.5	34615	1	BX569695 Synechoco
37	381.5	7.4	2801	5	AF218784 Gallus ga
38	380	7.4	209091	10	AC122299 Mus muscu
39	372.5	7.3	2015	8	CREVSPJA Chlamydomon
40	370	7.2	206849	2	AC011760 Drosophil
41	369.5	7.2	149159	10	AC148329 Mus muscu
42	368	7.2	242587	10	AC093481 Mus muscu
43	367.5	7.2	215910	10	AC127335 Mus muscu
44	367	7.1	7085	2	AC015260 Drosophil
45	367	7.1	179386	3	AC104515 Drosophil

#### ALIGNMENTS

RESULT 1	AX700058	2869 bp	DNA	linear	PAT 03-APR-2003
LOCUS	Sequence 2 from Patent WO03012090.				
DEFINITION	AX700058				
ACCESSION	AX700058				
VERSION	AX700058.1	GI:29536021			
KEYWORDS	Acidothermus cellulolyticus				
SOURCE	Acidothermus cellulolyticus				
ORGANISM	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Frankineae; Acidothermaceae; Acidothermus.				
REFERENCE	1				
AUTHORS	Ding, S.Y., Adney, W.S., Vinzant, T.B. and Himmel, M.E.				
TITLE	Thermal tolerant avicelase from Acidothermus cellulolyticus				
JOURNAL	Patent: WO 03012090-A 2 13-FEB-2003;				
FEATURES	Midwest Research Institute (US)				
source	Location/Qualifiers				
	1. .2869				
	/organism="Acidothermus cellulolyticus"				
	/mol_type="unassigned DNA"				
	/db_xref="taxon:28049"				

#### ORIGIN

Alignment Scores:	1.14e-146	Length:	2869
Pred. No.:	5134.00	Matches:	956
Score:	100.00%	Conservative:	0
Percent Similarity:	100.00%	Mismatches:	0
Best Local Similarity:	100.00%		



Query Match:	99.98%	Indels:	0
DB:	6	Gaps:	0
US-09-917-376-1 (1-957) x AX700058 (1-2869)			
Qy	1	MetAspArgSerGluAsnIleArgLeuThrMetArgSerArgArgLeuValSerLeuLeu	20
Db	1	ATGGATCGTTCCGGAGAACATCCGCTCGACTATGAGATCAGACGATTTGGTATCACTGCTC	60
Qy	21	AlaAlaThrAlaSerPheAlaValAlaAlaAlaLeuGlyValLeuProIleAlaIleThr	40
Db	61	GCGCCCATCGTGGTGGTGGCGCGCGCTCGAGTTCGGCGATTCGCCCATCGCCATAACG	120
Qy	41	AlaSerProAlaHisAlaAlaThrThrGlnProIleThrTrpSerAsnValAlaIleGly	60
Db	121	GCITCTCTCGGACGCGCGGACGACCTCAGCCGTACCTGGAGCAACGTTGGCGATCGGG	180
Qy	61	GlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrVal	80
Db	181	GGCGCGCGCTTGTGACGGGATCGTCTTCAATGAAGGTGCACCGGGAATTTCTGTACGTG	240
Qy	81	ArgThrAspIleGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeu	100
Db	241	CGGACGACATCGGGGGATGATATCGATGGGATCGCCCAACGGGCGGTGGATCCCTCTT	300
Qy	101	LeuAspTrpValGlyTyrAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAsp	120
Db	301	CTGGATTGGGGGATGGAACTTGGGGGTACAAACGCGCTGTCAGCATTTGGCGACAC	360
Qy	121	ProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspPro	140
Db	361	CCGATCAATACTAACAGGTATGGCCCGCTCGGAATGTACACCAACAGCTGGGACCCA	420
Qy	141	AsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeu	160
Db	421	AACGACGAGCGATTCCTCGCTCGTCTGATCAGGCGCAACGTGGCAATAACCGCCCTG	480
Qy	161	ProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluAtrcLeuAlaValAsp	180
Db	481	CCGTTCAAGCTTGGCGGCAACATGCCGGGGTGGAAATGGGCGAGCGGCTTGGCGTGGAT	540
Qy	181	ProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSer	200
Db	541	CCAAACAATGACAACTTCGTATTTCCGGCCCGCGAGCGGCAAGGGCTCTGGAGAACG	600
Qy	201	ThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIle	220
Db	601	ACAGATTCCGGCGCGACCTGGTCCAGATGCGAACTTTCCGACGCTAGGCACGTCACATT	660
Qy	221	AlaAsnProThrAspThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAla	240
Db	661	GCAATCCCACTGACACACCGCGCTATCAGAGCGATATCAAGGCGCTCGTCTGGGTGCGT	720
Qy	241	PheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAla	260
Db	721	TTGACAAAGTCTTCTGTCATCGCTCGGCAAGCGAGTAAACCATTTTGTGGGCGTGGCG	780
Qy	261	AspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValPro	280
Db	781	GATCCCAATATCCGCTCTCTCGAGACAGAGACGCGCGCGACGCTGGCAGCGGTGGCG	840
Qy	281	GlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeu	300
Db	841	GGTGGCGGACCGGCTTCATCCGCAACAGGGCGTCTTTGACCCCGTCAACCCAGCTGCTC	900
Qy	301	TyrIleAlaThrSerAsnThrGlyProTyrAspGlySerSerGlyAspValTrpLys	320
Db	901	TATATTGCCACAGCAATACGGGTGGTCCGTATGACGGAGCTCCGGCGACGCTCTGGAAA	960
Qy	321	PheSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAspThrAla	340
Db	961	TTCTCGGTGACCTCCGGGACATGGAGCGGATCAGCCCGGTACCTTCGACGACACGGCC	1020
Qy	341	AsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMet	360
Db	1021	AACGACTACTTTGGTTACAGCGGCTCTACTATCGACCGCCAGCACCAGCAAGATAATG	1080
Qy	361	ValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGly	380
Db	1081	GTGGCAACCCAGATATCGTGGTGGCGGACACCAATAATCTTTCCGAGACCCAGCGCGGT	1140
Qy	381	AlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrVal	400
Db	1141	CGGAGCTGGACGCGGATCTGGGATTTGGACGAGTATCCCAATCGAAGCTTGGATATGTG	1200
Qy	401	LeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProProValPro	420
Db	1201	CTTGACATTTCCGCGGAGCCTTGGCTGACCTTCGGCGTACAGCCGAATCTCCTCCGTACCC	1260
Qy	421	SerProLysLeuGlyTyrMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg	440
Db	1261	AGTCCGAAGCTCGGCTGGATGATGAAGCGATGGCAATCGATCCGTTCAACTCTGATCGG	1320
Qy	441	MetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSer	460
Db	1321	ATGCTCTACGGAAACAGCGCGAGCTTTGTACGCAACAATGATCTCACGAAGTGGGACTCC	1380
Qy	461	GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAsp	480
Db	1381	GGCGSCCAGATTCATATCGCGCGGATGGTCAAGGATTTGGAGGAGCGCGGTAAACGAT	1440
Qy	481	LeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPhe	500
Db	1441	CTCATCAGCCCGCTCTGGCGCCCGCTCATCAGCGCTCTCGAGACCTTCGGCGGCTTC	1500
Qy	501	ThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThr	520
Db	1501	ACCCACGCGCGATTTACTGCCGTGCCATCGACGATCTTTCAGTCCCGGTGTTCAGACC	1560
Qy	521	GlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPhe	540
Db	1561	GGCACCGGCTCGACTATCGGAAATTTGAATCCGTGATCATCTTCGCGTGGAAAGTTTC	1620
Qy	541	AspProSerSerGlnProAsnAspArgHisAlaPheSerThrAspGlyGlyLysAsn	560
Db	1621	GATCCATCGAGCAACCCGAACAGCAGCGACGTCGCGCTTCTCGACAGCGCGCGCAAGAC	1680
Qy	561	TrpPheGlnGlySerGluProGlyValThrThrGlyGlyThrValAlaAlaSerAla	580
Db	1681	TGGTTCCAGGCGCGAACCTGGCGGGGTGACAGCGGGGCGCACCGTCCGCGCATCGGCC	1740
Qy	581	AspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaVal	600
Db	1741	GACGCTCTCGTTTCTGGGCTCCCGGGATCCCGGTCCAGCTGTGGTGTACGCAAGTC	1800
Qy	601	GlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSer	620
Db	1801	GGATTTGGCAACTCTCTGGGCTGTTCCGAAGGTGTTCCCGCAATGCCAGATCCCGTCA	1860
Qy	621	AspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThr	640
Db	1861	GACCGGGTGAATCCAAAGACTTTTCTATGCCCTATCCAAATGGAACCTTCTATCGAAGCAGC	1920
Qy	641	AspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGly	660
Db	1921	GACGGCGGCTGACATTCACCGCTCGCGCGCGGTCTTCCGAGCAGCGGTGCGTCCGT	1980
Qy	661	ValMetPheHisAlaValProGlyLysGlyAspLeuTrpLeuAlaAspSerGly	680
Db	1981	GTCTGTTCCACGCGGTGCTGGAAAAGAGCGATCTGTGGTCTGCTGCTGATCGAGCGGG	2040
Qy	681	LeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAla	700
Db	2041	CTTTACCACTCAACCAATGGCGGACGATGCTGTGCAATCAACCGCGGTATCTCCGCG	2100
Qy	701	ValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValVal	720







```

/db_xref="UniProt/TREMBL:Q9ZBJ2"
/translaction="MAQVQATEERITVAADAEPTFDTLADYSGTGREKLIPHEFSYEVR
EGDGEGLVHMKLOATSKRVDRDCLLEVTEDGELVEKDRNSMWITWRTVPAGEGK
SRVVLTTMOGAGGVGGFFKTFAPKGLARIYDALLAKLAEEVK"
2394..4841
/gene="SC06483"
/notes="synonym: SC9C7.19"
2394..4841
/gene="SC06483"
/notes="SC9C7.19, probable efflux protein, len: 815aa;
similar to many, several of which are involved in toxin
antibiotic efflux eg. TR:069070 (EMBL:AF061335)
oxytetracycline exporter from Streptomyces rimosus (536
aa) fasta scores; Opt: 1020, z-score: 879.9, E(): 0,
(36.9% identity in 542 aa overlap)."
/codon_start=1
/transl_table=11
/product="putative efflux protein"
/protein_id="CAA22731.1"
/db_xref="GI:4154078"
/db_xref="GOA:Q9ZBJ1"
/db_xref="UniProt/TREMBL:Q9ZBJ1"
/translaction="MGGITLVQDPVPQDEPVVAQAQPGPPVPGDPGRDAGLSP
RRVRLVFGMLALLAALAEQMI VATALPKIVGELGLDKMSWITAYLLSTVGLPI
YKGLDILFGKGVFAILVVFVGSALAGRAQTMQDLIAFRAVQGVGAGGLMIGVQAI
IADIVPRERARYMGLIGASVAGPLGLGYFTDHLNWRWCFYINVPGLVTVAV
VAVLKLPRPRKRLDILGALLAAASTCLVLTSMGTEYANGSRVILGLAAGAVA
ATLLFLVAHPAPELPLRLPRDSIFNVTALVGLVGVVALFGAASYLPTFLQWVDA
ATSSGLMLPMMGVGASIIISQLISRTGHRHPLGLSALSVVGMWLLSRIDADT
SRLYSIWMVAGIAGVPLVLAQNSRPTDGLTATSNANNYFRQIGSGVGAAP
TLTFAGRLTALADRI PAEAGVGLPDAEAITPQLVHSLPPALRDAYIGAYADAMPRI
LYLVPVLGLLIALFLKEKPLVSHNAPVTPETAHPAQPPAAQVPPAPSPVAGGI
PVCGTVQHPDGTVPRAALTJLDVGGQIGRSGASDDGRVALATPGSGAVVLIAGG
HQPQAVSVTVGERVVELDVLGAGRLAGSVNTADGSPVRDATVTLNHNHGEVATFR
SGREGGVITELVAGEYITLAAPAFRAALPFAALPQAAARETRQDVELAGGAVLRGTVR
ASGRVADARVATLLDRAAGNVVDLTITGPDGTFRFVDLSSGEYTVIAGYPPVATVQ
VAGGRTERDILQLGHED"
4911..7493
/gene="SC06484"
/notes="synonym: SC9C7.20"
4911..7493
/gene="SC06484"
/notes="SC9C7.20, conserved hypothetical protein, len:
860aa; similar to many hypothetical proteins egs.
TR:086875 (EMBL:U22894) from Streptomyces lividans (547
aa) fasta scores; opt: 1000, z-score: 738.2, E(): 0,
(40.0% identity in 497 aa overlap) and TR:086807
(EMBL:AL031031) from Streptomyces coelicolor (916 aa)
fasta scores; opt: 358, z-score: 316.9, E(): 2.3e-10,
(27.2% identity in 779 aa overlap). Also contains a short
region of similarity to TR:Q53897 (EMBL:X60316) AbaA
regulatory locus from Streptomyces coelicolor (192 aa)
fasta scores; opt: 108, z-score: 223.2, E(): 3.8e-05,
(38.5% identity in 135 aa overlap)."
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAA22732.1"
/db_xref="GI:4154079"
/db_xref="GOA:Q9ZBJ0"
/db_xref="UniProt/TREMBL:Q9ZBJ0"
/translaction="MVTGHPRTVVTATQILRSRQGRGPDAMDHGTERRDAPPCHRAGPD
TVPADPVTGVLVAVVDRLGVSHWSRGARRLFAUSKEEAI GRPAPDILLPVSGLVP
EDAPGADGHDDSVRSYTAHDGPGPGLESSLDRLSVPAAGRARI TADDRDVRVLMW
AYPLVGLPSSVGLADADGLRRDYADGTDSDAGPAVVERIAFAFALHTDPFGADEL
ARRLPETLPSMVGESARI VAQVLEGLPVLEFQSDHRVPVTPDWGVPPRAERARRE
RAARALAGEVPDELDEAELEVAARERLEFLNEVSGAIGTSLDLSRTIVESRA
VVPRTDVAGTYLREQVAVGEGFPDGTDTTMMHVALEHTDPEGRWDDVVPVGEAM
FPFAPHTFFQVLTREQVAVGEGFPDGTDTTMMHVALEHTDPEGRWDDVVPVGEAM
VVLGFMILLRRHPVPFNDMDRVTCAEALRAAGLVLNARNMYTHQENVAETLQDSMLP
HIPARLAGCDITATYLPGLTILRGVGGWFDVAVKLPARTALVVGDMGHGNSAAMWG
QLRTAVOTMAGLDLPPAQLLRNLDLQALRGDTHLATCLYAVYDP IAGELHLNAGHI
PPVLVRAEDGASELIDLPTGAPIGVGGVPFESVRVTVAPGRDLVNVCTDGLVEVRGEDI
```

```

GVGLATLCESAAHPAASWMDACDTTIIRALGAGGGRKDDVALLMARITGTEPDAVAEWR
LARDPAEYGRARAAREQLYDMGLPKVPAELMVSVELTVNVRSHSHAPVALRLVRA
DITLCEVDDDDHELPAKSGPDBETGRGLRVVSTLAREWGASRTGAGKTVWFELTP
ARHR"
7711..8433
/gene="SC06485"
/notes="synonym: SC9C7.21"
7711..8433
/gene="SC06485"
/notes="SC9C7.21, unknown, len: 240 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC9C7.21"
/protein_id="CAA22733.1"
/db_xref="GI:4154080"
/db_xref="GOA:Q9ZBJ9"
/db_xref="UniProt/TREMBL:Q9ZBJ9"
/translaction="MSVTSRYRDAMGEFWNEASGRGAVFWDAEPTLTAGRHIAHPEP
HLTDAALPLIDLGCNGGTQTRYFAERFPVVGADLSAALAHARHADPGCLAAYRQLD
ADKAEQAQALHAELGDANVYVRGVHLQCPEPDQRLADALAALLGERGAFIVPEQQA
AKTVLMQAQGPAGPPAKLAPVLRHGLTPGAVPDEAVPEYLRAGLTVLASGELPLVT
TEYTAGSTRIELPSAWVVGTA"
8603..9436
/gene="SC06486"
/notes="synonym: SC9C7.22"
8603..9436
/gene="SC06486"
/notes="SC9C7.22, possible transport associated protein,
len: 277aa; similar to SW:DPPA_BACSU dppa, dipeptide
transport protein (274 aa) fasta scores; opt: 448,
z-score: 604.3, E(): 2.3e-26, (32.3% identity in 269 aa
overlap)."
/codon_start=1
/transl_table=11
Alignment Scores:
Pred. No.: 3,448-65 Length: 296500
Score: 2493.00 Matches: 480
Percent Similarity: 67.33% Conservative: 132
Best Local Similarity: 52.81% Mismatches: 236
Query Match: 48.55% Indels: 61
DB: 1 Gaps: 18
US-09-917-376-1 (1-957) x SCO939128 (1-296500)
Qy 10 ThrMetArgSerArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaValAla 29
|||||
|||||
70562 ACCGTGCGAAGAACCCGGATCCTCACCGTGTGTGGCCCTCGCGCCGCTCTGCTGGCG 70503
Qy 30 AlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAla-----HisAlaAla 47
|||||
|||||
70502 -----GGCAGTCCGCCCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 70452
Qy 48 ThrThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyGlyPheValAspGly 67
|||||
|||||
70451 GCCGCCGACAGCTACACCTGGAAGAACGCCCGCATCGACGCGCGGTTCGTCCCGCGC 70392
Qy 68 lleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMet 87
|||||
|||||
70391 ATCGTCTTCAACCGGACCGGAGAACCTGGCGCTACGCCCGCACCATCGGAGCGGC 70332
Qy 88 TyrArgTTPAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTTPAsn 107
|||||
|||||
70331 TACCCTGGCAGGAGAGTCCACACCTGGAGCGCGCTCTCGACACCGCTGGCTGGGAC 70272
Qy 108 AsnTTPGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysVal 127
|||||
|||||
70271 GACTGGGGGACACCGGGGTGTGCTCTCGCTCCGACCGCGCTCGATCCGACCGCGGC 70212
Qy 128 TrpAlaAlaValGlyMetTyrThrAsnSerTTPAspProAsnAspGlyAlaIleLeuArg 147
|||||
|||||
70211 TACGGCGCGGTGGCACGTACACCAACGACTGGGACCCGCCACCAACGCGCGGTGCTGCGC 70152
```



148	SerSerAspGlnGlyValaThrTrpGlnIleThrProLeuProPheThyLeuGlyGlyAsn	167	69080	ACC	GAGGTCCCGTCGATGATGTACACTCGCCCAACTTACCTTCGACGACGAGCCTGGAC	69022
70151	TCGCGGACCGGGTTCGAGCTGGGAGAGCGGACCTGCCGTTCAAGCTGGGCGGCAAC	70092	Qy	526	TyTAlaGluLeuAenProSerIleIleValArgAlaGlySerPheAspProSerSerGln	545
168	MetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAenAenAspAsnIleLeu	187	Db	69020	TTCCGGAGAGCAAGCGGACGTCGTGTCGGGCGGCAATCTCGAC-----TCCGGT	68967
70091	ATCCCGCGCGCGCATGGGTGAGCGCTGGCGCTGCGACCGCGACGACGAGCTGCTG	70032	Db	546	ProAsnAspArgHisValAlaPheSerThrAspGlyGlyIysAsnTrpPheGlnGlySer	565
188	TyrPheGlyAlaProSerGlyIysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrp	207	Qy	68966	CCG-----CACATTCGCGTTCTCCACGACAAACGCGCCCAACTGTTTCGGCGGCACC	68916
70031	TACCTGGGCGGCCCCAGCGGGCCACGAGCTGTGGGGTCCACGACGGGGGCTCACCTGG	69972	Db	566	GluProGlyGlyValThrThrGlyIleValAlaAlaSerAlaAspGlySerArgPhe	585
208	SerGlnMetThrAsnPheProAspValGlyThrTyrlleAlaAenProThrAspThrThr	227	Qy	68915	GACCCGTCGGGGTGAGCGCGCGCACGCTCGCGCGCGCGGCGGACGCGGCGCTTC	68856
69971	TCGAGGTGACGGCCCTTCCGNAACCCCGGGAACCTAGCGCAGGACCCGAAACACGTC	69912	Db	586	ValTrpAlaProGlyAspProGlyGlnProValTyrlleAlaValGlyPheGlyAsnSer	605
228	GlyTyrlleGlnSerAspIleGlnGlyValTrpValAlaPheAspLysSerSer---Ser	246	Qy	68955	GTGTGGAGCCCG-----GAGGGTCGCGGTCGTCGATACGACCGGCTTCGTTACGTC	68802
69911	CGCTACGCTCCGACCAACAGCGGCATCACCTGGGTTCACCTTCGACGAGTCGACGGG	69852	Db	606	TrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnPro	625
247	SerLeuGlyGlnAlaSerIleThylePheValGlyValAlaAenProAenAsnProVal	266	Qy	68801	TGGCAGCGCTCGACCGCGCTCCCGCGCGGCGCATCTCGTAGTCGACCGGCGTAACCC	68742
69851	GCGCGGGCAGCGCCACGCGACCTCTACGTCGGGGTCGGGGAACAGGAGACCGCGTC	69792	Db	626	LysThrPheTyrlleAlaLeuSerAsnGlyThrPheTyrlleArgSerThrAspGlyValThr	645
267	PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe	286	Qy	68741	CGACACCTTCTACGCTTCAAGTCCGGGAGGTTCCTACGTCAGTACGAGCGCGCGCAC	68682
69791	TACCCCTCGACCGACCGGGCGGACCTGGGAGCGGCTCGCGGGCAGCGACGGGTTAC	69732	Db	646	PheGlnProValAlaAla---GlyLeuProSerSerGlyAlaValGlyValMetPheHis	664
287	IleProHisLysGlyValPheAspProValAenHisValLeuTyrlleAlaThrSerAen	306	Qy	68681	TTTACCCCTCGCGCGCGACCGGGCTCGCGCGCGTGC-----GGCGTCCGCTTCAAG	68628
69731	CTGGCCCAACAGGGTGTCTGACCGCGGAGAACGGCTACCTGTACTCGCTACAGCGAC	69672	Db	665	AlaValProGlyIysGluGlyAspLeuTrpLeuAlaAlaSerSer-----	679
307	ThrGlyGlyProTyrlleAspGlySerSerGlyAspValTrpLysPheSerValThrSerGly	326	Qy	68627	GGCTGCGCGCGGGAGGGCGACGCTCTGGCTGGCGCGCGCGGCGGACGCGCCGTAC	68568
69671	ACCGCGCGCCCTACGACCGCGGCAAGGGCGGCTGTATCCCGGTACGCGCGGACCGGC	69612	Db	680	GlyLeuTyrlleHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSer	699
327	ThrTrpThrArgIleSerProValProSerThrAspThrAlaAenAspTyrllePheGlyTyrlle	346	Qy	68567	GGCTGTGGCACTCGACGCGCGCGCGGACCTTTCACAGGCTGCCCGGTGTGACGCGG	68508
69611	ACCTGACGACATCAGCCCGCGCGCGGCGGACAC-----TACTACGGCTTC	69561	Db	700	AlaValAenValGlyPheGlyLysSerAlaProGlySerSerTyrlleProAlaValPheVal	719
347	SerGlyLeuThrIleAspArgGlnHisProAenThrIleMetValAlaThrGlnIleSer	366	Qy	68507	CGGACACCGTTCGCTTCGGCAGCGCGCCCGCGGAGGGTGTACACGACCTTCTCAC	68448
69560	AGCGGCTCAGCTCGACCGCGCAGCGCGCGGCGCGGTGATGGCGACCGCTACAGCTCG	69501	Db	720	ValGlyThrIleGlyValThrGlyAlaTyrlleArgSerAspAspCysGlyThrTrp	739
367	TrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgile	386	Qy	68447	AGCGCGGAGATCGCGGGTGCAGCGCATCTTCGCTCCACCGACGCGGGCGGACCTGG	68388
69500	TGGTGGCGCGACACGAGATCTTCGCTCGACGGAACGCGCGCGACATGTGTCCCGCG	69441	Db	740	ValLeuIleAenAspAspGlnHisGlnTyrlleGlyAsnTrpGlyGlnAlaIleThrGlyAsp	759
387	TrpAspTrpThrSerTyrlleProAenArgSerLeuArgTyrlleValLeuAspIleSerAlaGlu	406	Qy	68387	ACCGGGTTCACGACGACGCGCCACGATGGGGCTGGACCGCGCGCGCATCACCGGAC	68328
69440	TGGAGCTACCTCTGTACCGGACCGCGGACCGGACCGCTACACGATGAGCTCTCTCGTCC	69381	Db	760	HisAlaAenLeuArgArgValTyrlleGlyThrAsnGlyArgGlyIleValTyrlleGly	778
407	ProTrpLeuThrPheGlyValGlnProAenProProValProSerProLysLeuGlyTrp	426	Qy	68327	CCGCGGTCTACGGCCGGGTGTAGTGGCGACCAACGCGCGCGGGGTGTACTACGGCGAC	68268
69380	CCCTGGCTCACCTGGGCGGCAACCGCGCGCGCGCGGACGACCCCGAAGCTGGGCTGG	69321	Db	779	-----AspIleGly-GlyAlaProSerGlySerProSerProSerValSerPro-----	794
427	MetAspGluAlaMetAlaIleAspProPheAenSerAspArgMetLeuTyrlleThrGly	446	Qy	68267	ACCTCGGACACGGCGGGGCGACCGATCCGGGTCCCGGACCGGACCCCGACCCCGACCG	68208
69320	ATGACGGAGGCTCTGGAGATGACACCGCTTCGACTCGGACCGGATGATGTACGGGACGGG	69261	Db	795	-----SerAlaSerProSerLeuSerProSerProSerProSerProSe	807
447	AlaThrLeuTyrlleAlaThrAenAspLeuThrLysTrp---AspSerGlyGlyGlnIleHis	465	Qy	68207	GCCTCGGAGGTGAGTACACGCTACGAAACAGTGGCGCGGGCGGCTTCCAGCGGACGTC	68148
69260	GCCACGGTCTACGGCACGGGAGAACCTGACGAACCTGGGACGACGAGGGCGGACGTTCCG	69201	Db	807	rSerSerProSerProSerProSerPro---SerSerSerProSerSerSerProSerPr	826
466	IleAlaProMetValLysGlyLeuGluGluThrAlaValAenAspLeuIleSerProPro	485	Qy	68147	CGCTTCAACACACCGGACGCTCGGCTTGGAAACGGGTGGTCCCTCGACTGGTCTTCCC	68088
69200	GTGAGCCGATGTGGCGGGCTTGGAGGAGAGCGGCGGTCAACGACCTGGCCTCCCGCG	69141	Db	826	o-----SerProSer---ProSerPro-----	832
486	SerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspVal	505	Qy	68087	GGCGGCCAGGAGGTGACCCCGATGTGGAAACCGCCGAGCACACCCAGCGCGGTACGTCG	68028



```

Db      68027  ACGGCGAGGAACGTGGCTGGGAACCGCGGTGTGGCCCCCGGTGCTCGCTTCGGG 67968
Qy      844  oSerProSerSerProSerProSerSerProSerSerProSerProSerProThPr 864
Db      67967  TTCACCGGAGCGCGTGGGACCGACCGGGAACCGGAGGGGTTCGCCGTCCCGGAGCG 67908
Qy      864  oSerSerProValSerGlyGly 872
Db      67907  GCGTCCCGGACCGCTACGTGACGGG 67883

RESULT 3
LOCUS   AP005028                      299800 bp    DNA    linear    BCT 10-MAY-2003
DEFINITION   Streptomyces avermitilis genomic DNA, complete genome, section 8/30
ACCESSION   AP005028  BA000030
VERSION     AP005028.1  GI:29605366
KEYWORDS
SOURCE      Streptomyces avermitilis MA-4680
ORGANISM    Streptomyces avermitilis MA-4680
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE
AUTHORS    Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Hanamoto,A., Takahashi,C., Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T., Kikuchi,H., Shiba,T., Sakaki,Y. and Hattori,M.
TITLE       Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
MEDLINE    21477403
PUBMED     11572948

REFERENCE
AUTHORS    Ikeda,H., Ishikawa,J., Hanamoto,A., Shinose,M., Kikuchi,H., Shiba,T., Sakaki,Y., Hattori,M. and Omura,S.
TITLE       Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis
JOURNAL     Nat. Biotechnol. 21 (5), 526-531 (2003)
MEDLINE    22608306
PUBMED     12692562

REFERENCE
AUTHORS    Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C., Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T., Kushida,N., Director-General of Biotechnology Center, Shiba,T., Sakaki,Y. and Hattori,M.
TITLE       Direct Submission
JOURNAL     Submitted (29-MAR-2002) Director-General of Biotechnology Center, National Institute of Technology and Evaluation, Biotechnology Center; 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan (E-mail:bio@nitech.go.jp, URL:http://www.bio.nitech.go.jp/, Tel:81-3-3481-1933, Fax:81-3-3481-8424)
COMMENT     This work was done in collaboration with Haruo Ikeda(*1), Jun Ishikawa(*2), Akiharu Horikawa(*3), Chigusa Takahashi(*3), Mayumi Shinose(*3), Hiroshi Horikawa(*4), Hidekazu Nakazawa(*4), Tomomi Osonoe(*4), Norihiro Kushida(*4), Hisashi Kikuchi(*4), Tadayoshi Shiba(*5), Yoshiyuki Sakaki(*6,*7), Masahira Hattori(*1,*7) and Satsoshi Omura(*1,*3).
FINAL FINISHING PROCESS AND ALL ANNOTATION WERE DONE BY H. IKEDA AND J. ISHIKAWA.
*1 Kitasato Institute for Life Sciences, Kitasato University
*2 National Institute of Infectious Diseases
*3 The Kitasato Institute
*4 National Institute of Technology and Evaluation
*5 School of Science, Kitasato University
*6 Institute of Medical Science, University of Tokyo
*7 RIKEN, Genomic Sciences Center
Following url is also available.
http://avermitilis.is.kitasato-u.ac.jp.
Location/Qualifiers
1. .299800
/organism="Streptomyces avermitilis MA-4680"
/mol_type="genomic DNA"
/strain="MA-4680"

FEATURES
source

```

```

/db_xref="taxon:27882"
/note="This strain is also named as strain: ATCC 31267, NCIM 12804 or NRRL 8165.-synonym: Streptomyces avermectinius"
16. .918
/note="SAV1722"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC69433.1"
/db_xref="GI:29605367"
/translation="MAGRALAVPCADLTPLMTSAESDGPGRVDRLLRPVPSKMLGEE LPAVEVTERGLAHDRRVALIDRDITGKVASAKHPLRWLLITVTAQADNSAVRMTSDG KELWSTDGVDSEISEIVGRPVFLTDPAPHEALDRADPEDVLRDTRPPVSARTGRI KGSSEGTFFDFAPVHMTTSLARIALAPNGSAQAERYRNPIVLRPLTRGTEND WIDRTLHGTLGLVLRVIARTPRCAIPTLAHGTVPADTAALRLVLAENHVRVTPTESMGPG PCAGYVAQVLRPGRIRLGDPPVHA"
complement(1232..3406)
/note="SAV1723"
/codon_start=1
/transl_table=11
/product="putative mycodextranase"
/protein_id="BAC69434.1"
/db_xref="GI:29605368"
/translation="MHSNLSRYRRMSAIGAVALAAGLLVALAPAAHAAAGADLPFT PVQESATTTGKIGPDHTQGTIASEASQAVRLGAGQRTVEFTVPRAANAVNVSYPG SDQSGALNVYNGTKLAKTLAVTSKYSYVDTSWIAAGQTHFFDNARLLGQNVPG DKLIAEAVSTQVVDVADPEOVAARTKPAVASVTDKGADSGQSDGQAERNALAA AQGVVWIPPGDYRLTSLINGQNTVLQAGSWSHVHTSRFIDQSSSGGVHLKQFA VIGEVTERVDNPNFVNGSLGSSVSGMIQHLKVGLWLTGNNDLVVENNRILDT TADGLNLNAGRVRRNFRNNGD DALAMSLYSPDNTSFESNTILVFRAGAMPN IYGGTDLAVKXNLI SDTNALSGSIAI SNOKFLDPFSPLAGTITVSGNTILVFRAGAMPN WHPMGLRVDSYDSAI BANYSTIGTTITDSYSAFEFVSGGCHGYATKNVTVDGATV RNTGTVVQAESQGAKEFNQATGVGAGI VNCYPAGSGTFTLTDGGNAGWSTW GDCSWPQGGNPDPTPRNLAKGRPATATGSQDYVTECKAVDGDASTYWESTNRAP PQATVTDIGSSQAVRLVLPPTAPMAQARTQTLSVQSGTDSGASVTVVASQGYRDFP TGTNTATVTPSGTNRLYLKLNVTANTGTGMPAAQGFSEVEAYLS"
complement(3464..5065)
/gene="aglA1"
complement(3464..5065)
/gene="aglA1"
/note="SAV1724"
/codon_start=1
/transl_table=11
/product="putative alpha-glucosidase"
/protein_id="BAC69435.1"
/db_xref="GI:29605369"
/translation="MGQPTPARIQKDMWRMTAVIQQYVRSFADGDGDTGLAGVRA R LPYLAGLVDAIWFNPWYLS PMKGGYDVADYRVPDPAPFTGLAEAEKLI AEARELGIR TLVDIVPNHVSQHPFRAALAAAGPGSPERALPHFRGGRGAGELPNDWPSPQFAGT R EVMTQLPDGDWYHLFTPEQDLNWAHPVQEHEDVLRFWFERGVAGVRI DSAALL AKDPDLQDVEGRAPHPFVDRDELHDIRSWRAIADYGAVFGEVWLPDTERFAYL RDETLDFNFSLCSQPDMDARLRTAIDTAEHAPVGPATVWLNCHDVIRTVIRYG RTDTGDFAAKAGFTTDLALGTRRAAALISLALPGSVLYLQSGELGPEAEI PRE STQDPMHFRSHGVDPGRDCRVPVPAVADHAGFGGEPWLPQPARWSAYAADLQAE D PGMSLSLYREALRLRPVADGPLTWLPADGVLAFTARTGLMCMVNLADTTPAGLPAHS RLLLSGSLDQGLPQDTAAWLRI"
complement(5097..5972)
/note="SAV1725"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="BAC69436.1"
/db_xref="GI:29605370"
/translation="MSTRTLVSPSLTSLRPRGRKAAVWTVFTGVLLFALLAFVPPYMWV TGA M KSPDEARTPPSLVQKWLSGYTDWDLMLQPLAHLWNTVTVADPLIHTSLLDVLCI TAAAYALSRLKPVFGKLVLGILATLMVPAQALVVKPLTVADPLIHTSLLDVLCI WLPVANAFLYLLKRFDDQIPRDVLEAAEIDGAGRLFTLSIVLEPMSRLGVSTIF ALPVAWQDFLWPLVMFVSTDQKQFISVALVQLSQNLQTLTVLIAAWIASIPMWAMFLVF QRHIIAGISAGSTKG"
complement(5969..6925)
/note="SAV1726"
/codon_start=1

```



CDS	/transl_table=11		DB:	1	Gaps:	18
	/product="putative integral membrane protein"		US-09-917-376-1 (1-957) x AP005028 (1-299800)			
	/protein_id="BAC69437.1"		QY	9	LeuThrMetArgSerArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaVal 28	
	/db_xref="GI:29605371"		DB	175638	CTCACGTCGGAAGAACCCGCATCTTC-----ACGGCGGTGCTCGCGCTG 175682	
	/transl_table=11		QY	29	AlaAlaAlaLeu-----GlyValLeuProIleAlaIleThrAlaSerProAlaHisAla 46	
	/note="SAV1727"		DB	175683	CGCGCGGTCTGCCCGCGGCACCCCGCGGCTGCGCGGAGCGGCCCGCGGACG 175742	
	/codon_start=1		QY	47	AlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyGlyPheValAsp 66	
	/product="putative sugar-binding lipoprotein precursor"		DB	175743	ATCCCGCGGACACGTACAGTCGAAAGACGCCCGCGCTGCGCGGCGGCTTCGTCCCC 175802	
	/transl_table=11		QY	67	GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGly 86	
	/db_xref="GI:29605372"		DB	175803	GGCATCGTCTTCAACCGCTCCGAGAAAGACCTCGCTTACGCCCGCACCGACATCGGCGGC 175862	
	/transl_table=11		QY	87	MetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTyr 106	
	/note="SAV1728"		DB	175863	GCCTACCGCTGGCGCCGAGTCTCTGAAGACCTGAGCGCGCTGCTCGACTCGTTCGGCTGG 175922	
	/codon_start=1		QY	107	AsnAsnTrpGlyTyrAsnGlyValSerIleAlaAlaAspProIleAsnThrAsnLys 126	
	/product="putative LacI-family transcriptional regulator"		DB	175923	AGCGACTGGGGGCACACGGGTGCTGAGGCTCGCTCGACTCGCTCGACCGCAACG 175982	
	/transl_table=11		QY	127	ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 146	
	/db_xref="GI:29605373"		DB	175983	GTGTACCGCGCGCTCGGCACGTACAGAACAGCTGGGACCGCGGCAACGGTGCCTGCTC 176042	
	/transl_table=11		QY	147	ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGly 166	
	/note="SAV1729"		DB	176043	AGGTCCCGGCACCGGGCGCGAGCTGCGAAGAACCGACCTGCGCTTCAAGCTGCGCGGG 176102	
	/codon_start=1		QY	167	AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 186	
	/product="putative GntR-family transcriptional regulator"		DB	176103	AACATCGCGCGCGGGCATGGCGGAGCGGCTCGCGTTCGACCGCGAACAGACGCGTG 176162	
	/transl_table=11		QY	187	LeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThr 206	
	/db_xref="GI:29605374"		DB	176163	CTGTACTCTCGCGCGCCACGCGGCGTGTGGCGGTTCGACCGACTCGGGGCGCTCC 176222	
	/transl_table=11		QY	207	TrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 226	
	/note="SAV1730"		DB	176223	TGGTCGAGGTCCACCGACTTCCCGAACCTCGGCACTACGTCGAGGACCGCGACGACG 176282	
	/codon_start=1		QY	227	ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSer 246	
	/product="putative proline racemase"		DB	176283	AGCGGGTACGCGTCCGACAAACCGGGCATCGTGTGGGTTCACCTTCGACGAGTCGACGCGGG 176342	
	/transl_table=11		QY	247	SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAlaAspProAsnAsnProVal 266	
	/db_xref="GI:29605375"		DB	176343	TCGCGGGGAGCTCCACGCGGACGGTGTACGTGCGGGTTCGCGACAGGACACATCTCCGCTC 176402	
	/transl_table=11		QY	267	PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 286	
	/note="SAV1731"		DB	176403	TATCGCTCCACGAGCGCGCGCGACCTGTGTGCGCGCGCGCGCGCGCGCGCGCAT 176462	
	/codon_start=1		QY	287	IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn 306	
	/product="putative integral membrane protein"		DB	176463	CTCGCCCCAACAAAGGCGGTGCTGGACCGCGGCGAAACGGCTGTCTGTACTCTCGCTACGCGCAG 176522	
	/transl_table=11		QY	307	ThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGly 326	
	/db_xref="GI:29605376"		DB	176523	AAGGGCGGACCGTACGACGCGCGGCAAGGACAGTGTGGCGGTTCACCGCAAGACCGCGG 176582	
	/transl_table=11		QY	327	ThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr 346	
	/note="SAV1732"		DB	176583	ACCTGGACGAACATCAGCCCGGTTCGCGAGGCGCGACACC-----TACTACGGCTTC 176633	
	/codon_start=1		QY	347	SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer 366	

CDS	/transl_table=11		DB:	1	Gaps:	18
	/product="putative integral membrane protein"		US-09-917-376-1 (1-957) x AP005028 (1-299800)			
	/protein_id="BAC69437.1"		QY	9	LeuThrMetArgSerArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaVal 28	
	/db_xref="GI:29605371"		DB	175638	CTCACGTCGGAAGAACCCGCATCTTC-----ACGGCGGTGCTCGCGCTG 175682	
	/transl_table=11		QY	29	AlaAlaAlaLeu-----GlyValLeuProIleAlaIleThrAlaSerProAlaHisAla 46	
	/note="SAV1727"		DB	175683	CGCGCGGTCTGCCCGCGGCACCCCGCGGCTGCGCGGAGCGGCCCGCGGACG 175742	
	/codon_start=1		QY	47	AlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyGlyPheValAsp 66	
	/product="putative sugar-binding lipoprotein precursor"		DB	175743	ATCCCGCGGACACGTACAGTCGAAAGACGCCCGCGCTGCGCGGCGGCTTCGTCCCC 175802	
	/transl_table=11		QY	67	GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGly 86	
	/db_xref="GI:29605372"		DB	175803	GGCATCGTCTTCAACCGCTCCGAGAAAGACCTCGCTTACGCCCGCACCGACATCGGCGGC 175862	
	/transl_table=11		QY	87	MetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTyr 106	
	/note="SAV1728"		DB	175863	GCCTACCGCTGGCGCCGAGTCTCTGAAGACCTGAGCGCGCTGCTCGACTCGTTCGGCTGG 175922	
	/codon_start=1		QY	107	AsnAsnTrpGlyTyrAsnGlyValSerIleAlaAlaAspProIleAsnThrAsnLys 126	
	/product="putative LacI-family transcriptional regulator"		DB	175923	AGCGACTGGGGGCACACGGGTGCTGAGGCTCGCTCGACTCGCTCGACCGCAACG 175982	
	/transl_table=11		QY	127	ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 146	
	/db_xref="GI:29605373"		DB	175983	GTGTACCGCGCGCTCGGCACGTACAGAACAGCTGGGACCGCGGCAACGGTGCCTGCTC 176042	
	/transl_table=11		QY	147	ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGly 166	
	/note="SAV1729"		DB	176043	AGGTCCCGGCACCGGGCGCGAGCTGCGAAGAACCGACCTGCGCTTCAAGCTGCGCGGG 176102	
	/codon_start=1		QY	167	AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 186	
	/product="putative GntR-family transcriptional regulator"		DB	176103	AACATCGCGCGCGGGCATGGCGGAGCGGCTCGCGTTCGACCGCGAACAGACGCGTG 176162	
	/transl_table=11		QY	187	LeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThr 206	
	/db_xref="GI:29605374"		DB	176163	CTGTACTCTCGCGCGCCACGCGGCGTGTGGCGGTTCGACCGACTCGGGGCGCTCC 176222	
	/transl_table=11		QY	207	TrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 226	
	/note="SAV1730"		DB	176223	TGGTCGAGGTCCACCGACTTCCCGAACCTCGGCACTACGTCGAGGACCGCGACGACG 176282	
	/codon_start=1		QY	227	ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSer 246	
	/product="putative proline racemase"		DB	176283	AGCGGGTACGCGTCCGACAAACCGGGCATCGTGTGGGTTCACCTTCGACGAGTCGACGCGGG 176342	
	/transl_table=11		QY	247	SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAlaAspProAsnAsnProVal 266	
	/db_xref="GI:29605375"		DB	176343	TCGCGGGGAGCTCCACGCGGACGGTGTACGTGCGGGTTCGCGACAGGACACATCTCCGCTC 176402	
	/transl_table=11		QY	267	PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 286	
	/note="SAV1731"		DB	176403	TATCGCTCCACGAGCGCGCGCGACCTGTGTGCGCGCGCGCGCGCGCGCGCGCAT 176462	
	/codon_start=1		QY	287	IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn 306	
	/product="putative integral membrane protein"		DB	176463	CTCGCCCCAACAAAGGCGGTGCTGGACCGCGGCGAAACGGCTGTCTGTACTCTCGCTACGCGCAG 176522	
	/transl_table=11		QY	307	ThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGly 326	
	/db_xref="GI:29605376"		DB	176523	AAGGGCGGACCGTACGACGCGCGGCAAGGACAGTGTGGCGGTTCACCGCAAGACCGCGG 176582	
	/transl_table=11		QY	327	ThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr 346	
	/note="SAV1732"		DB	176583	ACCTGGACGAACATCAGCCCGGTTCGCGAGGCGCGACACC-----TACTACGGCTTC 176633	
	/codon_start=1		QY	347	SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer 366	

CDS	/transl_table=11		DB:	1	Gaps:	18
	/product="putative integral membrane protein"		US-09-917-376-1 (1-957) x AP005028 (1-299800)			
	/protein_id="BAC69437.1"		QY	9	LeuThrMetArgSerArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaVal 28	
	/db_xref="GI:29605371"		DB	175638	CTCACGTCGGAAGAACCCGCATCTTC-----ACGGCGGTGCTCGCGCTG 175682	
	/transl_table=11		QY	29	AlaAlaAlaLeu-----GlyValLeuProIleAlaIleThrAlaSerProAlaHisAla 46	
	/note="SAV1727"		DB	175683	CGCGCGGTCTGCCCGCGGCACCCCGCGGCTGCGCGGAGCGGCCCGCGGACG 175742	
	/codon_start=1		QY	47	AlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyGlyPheValAsp 66	
	/product="putative sugar-binding lipoprotein precursor"		DB	175743	ATCCCGCGGACACGTACAGTCGAAAGACGCCCGCGCTGCGCGGCGGCTTCGTCCCC 175802	
	/transl_table=11		QY	67	GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGly 86	
	/db_xref="GI:29605372"		DB	175803	GGCATCGTCTTCAACCGCTCCGAGAAAGACCTCGCTTACGCCCGCACCGACATCGGCGGC 175862	
	/transl_table=11		QY	87	MetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTyr 106	
	/note="SAV1728"		DB	175863	GCCTACCGCTGGCGCCGAGTCTCTGAAGACCTGAGCGCGCTGCTCGACTCGTTCGGCTGG 175922	
	/codon_start=1		QY	107	AsnAsnTrpGlyTyrAsnGlyValSerIleAlaAlaAspProIleAsnThrAsnLys 126	
	/product="putative LacI-family transcriptional regulator"		DB	175923	AGCGACTGGGGGCACACGGGTGCTGAGGCTCGCTCGACTCGCTCGACCGCAACG 175982	
	/transl_table=11		QY	127	ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 146	
	/db_xref="GI:29605373"		DB	175983	GTGTACCGCGCGCTCGGCACGTACAGAACAGCTGGGACCGCGGCAACGGTGCCTGCTC 176042	
	/transl_table=11		QY	147	ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGly 166	
	/note="SAV1729"		DB	176043	AGGTCCCGGCACCGGGCGCGAGCTGCGAAGAACCGACCTGCGCTTCAAGCTGCGCGGG 176102	
	/codon_start=1		QY	167	AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 186	
	/product="putative GntR-family transcriptional regulator"		DB	176103	AACATCGCGCGCGGGCATGGCGGAGCGGCTCGCGTTCGACCGCGAACAGACGCGTG 176162	
	/transl_table=11		QY	187	LeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThr 206	
	/db_xref="GI:29605374"		DB	176163	CTGTACTCTCGCGCGCCACGCGGCGTGTGGCGGTTCGACCGACTCGGGGCGCTCC 176222	
	/transl_table=11		QY	207	TrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 226	
	/note="SAV1730"		DB	176223	TGGTCGAGGTCCACCGACTTCCCGAACCTCGGCACTACGTCGAGGACCGCGACGACG 176282	
	/codon_start=1		QY	227	ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSer 246	
	/product="putative proline racemase"		DB	176283	AGCGGGTACGCGTCCGACAAACCGGGCATCGTGTGGGTTCACCTTCGACGAGTCGACGCGGG 176342	
	/transl_table=11		QY	247	SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAlaAspProAsnAsnProVal 266	
	/db_xref="GI:29605375"		DB	176343	TCGCGGGGAGCTCCACGCGGACGGTGTACGTGCGGGTTCGCGACAGGACACATCTCCGCTC 176402	
	/transl_table=11		QY	267	PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 286	
	/note="SAV1731"		DB	176403	TATCGCTCCACGAGCGCGCGCGACCTGTGTGCGCGCGCGCGCGCGCGCGCGCAT 176462	
	/codon_start=1		QY	287	IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn 306	
	/product="putative integral membrane protein"		DB	176463	CTCGCCCCAACAAAGGCGGTGCTGGACCGCGGCGAAACGGCTGTCTGTACTCTCGCTACGCGCAG 176522	
	/transl_table=11		QY	307	ThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGly 326	
	/db_xref="GI:29605376"		DB	176523	AAGGGCGGACCGTACGACGCGCGGCAAGGACAGTGTGGCGGTTCACCGCAAGACCGCGG 176582	
	/transl_table=11		QY	327	ThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr 346	
	/note="SAV1732"		DB	176583	ACCTGGACGAACATCAGCCCGGTTCGCGAGGCGCGACACC-----TACTACGGCTTC 176633	
	/codon_start=1		QY	347	SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer 366	

CDS	/transl_table=11		DB:	1	Gaps:	18
	/product="putative integral membrane protein"		US-09-917-376-1 (1-957) x AP005028 (1-299800)			
	/protein_id="BAC69437.1"		QY	9	LeuThrMetArgSerArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaVal 28	
	/db_xref="GI:29605371"		DB	175638	CTCACGTCGGAAGAACCCGCATCTTC-----ACGGCGGTGCTCGCGCTG 175682	
	/transl_table=11		QY	29	AlaAlaAlaLeu-----GlyValLeuProIleAlaIleThrAlaSerProAlaHisAla 46	
	/note="SAV1727"		DB	175683	CGCGCGGTCTGCCCGCGGCACCCCGCGGCTGCGCGGAGCGGCCCGCGGACG 175742	
	/codon_start=1		QY	47	AlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyGlyPheValAsp 66	
	/product="putative sugar-binding lipoprotein precursor"		DB	175743	ATCCCGCGGACACGTACAGTCGAAAGACGCCCGCGCTGCGCGGCGGCTTCGTCCCC 175802	
	/transl_table=11		QY	67	GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGly 86	
	/db_xref="GI:29605372"		DB	175803	GGCATCGTCTTCAACCGCTCCGAGAAAGACCTCGCTTACGCCCGCACCGACATCGGCGGC 175862	
	/transl_table=11		QY	87	MetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTyr 106	
	/note="SAV1728"		DB	175863	GCCTACCGCTGGCGCCGAGTCTCTGAAGACCTGAGCGCGCTGCTCGACTCGTTCGGCTGG 175922	
	/codon_start=1		QY	107	AsnAsnTrpGlyTyrAsnGlyValSerIleAlaAlaAspProIleAsnThrAsnLys 126	
	/product="putative LacI-family transcriptional regulator"		DB	175923	AGCGACTGGGGGCACACGGGTGCTGAGGCTCGCTCGACTCGCTCGACCGCAACG 175982	
	/transl_table=11		QY	127	ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 146	
	/db_xref="GI:29605373"		DB	175983	GTGTACCGCGCGCTCGGCACGTACAGAACAGCTGGGACCGCGGCAACGGTGCCTGCTC 176042	
	/transl_table=11		QY	147	ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGly 166	
	/note="SAV1729"		DB	176043	AGGTCCCGGCACCGGGCGCGAGCTGCGAAGAACCGACCTGCGCTTCAAGCTGCGCGGG 176102	
	/codon_start=1		QY	167	AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 186	
	/product="putative GntR-family transcriptional regulator"		DB	176103	AACATCGCGCGCGGGCATGGCGGAGCGGCTCGCGTTCGACCGCGAACAGACGCGTG 176162	
	/transl_table=11		QY	187	LeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThr 206	
	/db_xref="GI:29605374"		DB	176163	CTGTACTCTCGCGCGCCACGCGGCGTGTGGCGGTTCGACCGACTCGGGGCGCTCC 176222	
	/transl_table=11		QY	207	TrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 226	
	/note="SAV1730"		DB	176223	TGGTCGAGGTCCACCGACTTCCCGAACCTCGGCACTACGTCGAGGACCGCGACGACG 176282	
	/codon_start=1		QY	227	ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSer 246	
	/product="putative proline racemase"		DB	176283	AGCGGGTACGCGTCCGACAAACCGGGCATCGTGTGGGTTCACCTTCGACGAGTCGACGCGGG 176342	
	/transl_table=11		QY	247	SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAlaAspProAsnAsnProVal 266	
	/db_xref="GI:29605375"		DB	176343	TCGCGGGGAGCTCCACGCGGACGGTGTACGTGCGGGTTCGCGACAGGACACATCTCCGCTC 176402	
	/transl_table=11		QY	267	PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 286	
	/note="SAV1731"		DB	176403	TATCGCTCCACGAGCGCGCGCGACCTGTGTGCGCGCGCGCGCGCGCGCGCGCAT 176462	
	/codon_start=1		QY	287	IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn 306	
	/product="putative integral membrane protein"		DB	176463	CTCGCCCCAACAAAGGCGGTGCTGGACCGCGGCGAAACGGCTGTCTGTACTCTCGCTACGCGCAG 176522	
	/transl_table=11		QY	307	ThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGly 326	
	/db_xref="GI:29605376"		DB	176523	AAGGGCGGACCGTACGACGCGCGGCAAGGACAGTGTGGCGGTTCACCGCAAGACCGCGG 176582	
	/transl_table=11		QY	327	ThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr 346	

Alignment Scores:  
Pred. No.: 5, 27e-65 Length: 299800  
Score: 2487.00 Matches: 490  
Percent Similarity: 64.77% Conservativeness: 135  
Best Local Similarity: 50.78% Mismatches: 248  
Query Match: 48.43% Indels: 93











Qy 448 ThrLeuTyrAlaThrAsnAspLeuThrLysTyrAspSerGlyGlyGlnIleHisAla 467  
 Db 1567 AMATTGTATGATCGCATGTTGACAAATTTGGACAAAGGCAAAATATCACCAATAA 1626  
 Qy 468 ProMetValLysGlyLeuGluGluThrAlaValAsnAspLeuLysSerProSerGly 487  
 Db 1627 GTCAAGGCTATAGGAATTTAGAAATTTTCAAGCAAGCTGATAGGCAAGCAAGTGA 1686  
 Qy 488 AlaProLeuLeuSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAla 507  
 Db 1687 CCACATTTATTCAGCTCGGAGATATACGGGATTCAGACGAAGACTTCGGAGAA 1746  
 Qy 508 ValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAla 527  
 Db 1747 GCACCTAATTCGACATATTCACCTAATATGAGGACTACTACTGATATTTGCT 1806  
 Qy 528 GluLeuAsnProSerIleIleValAlaGlySerPheAspProSerSerGlnProAsn 547  
 Db 1807 GAGTTAAATCCAAATTTTATGTTTGTGAGGAAATGTTGATAAACAAGTGAATCCAAAT 1866  
 Qy 548 AspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGly---SerGlu 566  
 Db 1867 ACTAACAGAAATGTTTCTCATATGATGTTGGAATAATGTTGATAAACAAGTGAATCCAAAT 1926  
 Qy 567 ProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheVal 586  
 Db 1927 CCTCAAGAACCAAGGAGGTTGACAGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1986  
 Qy 587 TrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrp 606  
 Db 1987 TGGGCACCA-----AAAGGAGCGAAAGTATGCTATTCTACAGATAACGGAACAAATGG 2040  
 Qy 607 AlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLys 626  
 Db 2041 GTTCAATCGCAAAATGACCTTCAGAGCTATTGATATATCTGACAGAGTAAATCCAAAT 2100  
 Qy 627 ThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPhe 646  
 Db 2101 AAGTTTATGATTTAAGATGGCAAAATTTATATAGTGTGATGATAAAGGAAACATTT 2160  
 Qy 647 ---GlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAla 665  
 Db 2161 ATCGAAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2208  
 Qy 666 ValProGlyLysGluGlyAspLeuTrpLeuAlaLaserSerGlyLeuTyrHisSerThr 685  
 Db 2209 GTTCCAGGTATAGAGGTGATATCTGCTAGTGTGATGATGATGATGATGATGATGATGAT 2268  
 Qy 686 AsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPhe 705  
 Db 2269 GATGTTGATATTCATTTGTAATAATTTTCAAGGTTGAGGATGCGAGGATATAGGATTT 2328  
 Qy 706 GlyLysSerAlaProGlySerTyrProAlaValPheValValGlyThrIleGlyGly 725  
 Db 2329 GGTAAACCTGCTGAGGAGAAACATATCTGCAATTTATCTTATGCAAAATTAATGA 2388  
 Qy 726 ValThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAspAsp 745  
 Db 2389 GTAAAGAGTATATTAGATCTGATGATCTGATGATGATGATGATGATGATGATGATGATGAT 2448  
 Qy 746 GlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArg 765  
 Db 2449 AAACATCATGTTGGTTGTGCGAATGCGACACATACTGAGATCCACGGCTGTACGGTGA 2508  
 Qy 766 ValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyGlyAlaProSer 785  
 Db 2509 GTATTTGTTCGACAAACCGCTGGTATATAAAATGGGTGAATTTGCT-----TAT 2559  
 Qy 786 GlySerProSerProSerValSerProSerAlaSerProSerLeuSerProSerProSer 805  
 Db 2560 TCAATATTTCCGCATCAGCAACTCCACATCAACACCAAGTTTCAACCCCGACACCGACT 2619  
 Qy 806 ProSerSerSerProSerProSerProSerProSerProSerProSerProSerProSer 825

Db 2620 CCAACAGCAACTATTACATCCACACTGCACCAACATCAACCAACCAACCAACCCCGACC 2679  
 Qy 826 ProSerProSerProSerProSerProSerProSerProSerProSerProSerProSer 845  
 Db 2680 CCAACACCAACAGTACAGCAACCGCCGACGCAACCGCCGACGCTTACAGTACGCTTAT 2739  
 Qy 846 ProSerSerProSerProSerProSerProSerProSerProSerProSerProSerProSer 865  
 Db 2740 ATGACTCCGACACGACACCA-----ACACCAACCGCCGACACCAACA 2781  
 Qy 866 SerSerProValSerGly---GlyValLysValGlnTyrLysAsnAsnAspSerAlaPro 884  
 Db 2782 GGGACACCTGTTACGGGAAGTGGTTGAAGGTACTGTACAGCAACAAATGAGACGAGTGGC 2841  
 Qy 885 GlyAspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySerSerSerValAsp 904  
 Db 2842 AGTCAGGTTCTTATAAGGCCGCTGGTTAAGATAGTGAATGGAGGAGCAGCAGTGTGAT 2901  
 Qy 905 LeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuVal 924  
 Db 2902 CTTACAGGTTAAGATAAGTACTGTGTACACAGTGGATGGTACAGCAAGCAGAGTGGC 2961  
 Qy 925 TyrAsnCysAspTrpAlaAlaIleGlyCysGlyAsnIleArgAlaSerPheGlySerVal 944  
 Db 2962 GTA---TGTGACTGGCACAGATAGTGCAGCAATGTGACATTCAATTTTGTGAAGCTT 3018  
 Qy 945 AsnProAlaThrProThrAlaAspThrTyrLeuGln 956  
 Db 3019 AGCAGCGAGTGTGAGTGGAGCGGATTTACTTTGGAG 3054

RESULT 5  
 AX565635  
 LOCUS AX565635  
 DEFINITION Sequence 1 from Patent WO02077242.  
 ACCESSION AX565635  
 VERSION AX565635.1 GI:26000969  
 KEYWORDS  
 SOURCE unidentified  
 ORGANISM unclassified.  
 REFERENCE 1  
 AUTHORS Duffner, F. and Sjoeholm, C.  
 TITLE Family 74 xyloglucanases  
 JOURNAL Patent: WO 02077242-A 1 03-OCT-2002;  
 Novozymes A/S (DK)  
 FEATURES  
 Location/Qualifiers  
 1..2823  
 /organism="unidentified"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:3264"  
 /note="Jonesia sp., DSM 14140"  
 CDS  
 1..2823  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="CAD58415.1"  
 /db\_xref="GI:26000970"  
 /translation="MLRTVATATALLATTAGGIATSTPMVTPAPSPSSPSGYS  
 SWSNVEIVGGGFVPGIVFNQKDPGLVYARTIDIGAYRLNDSRTGRTPLTDHIGWDSY  
 HSGILSLATDPVTRNVLAAAGTSMWDQPNQAILRSADKGTWKTWMLPFRVGNM  
 PRCGMRGLAVDPNNKVLVFGAESGNGLWSTDKYKTKWKTSPFNAGNYVADASGA  
 YTGQVQVWVTFDPTSAKAGTTQTIYGVADKQNNVYRSTDDGATWQVPGQTGP  
 LAQGLVHLKQQLYATSDTGPYDGSKGDVWRDLISGGQWTRISPIESTSSNSAFG  
 YSGLAIDRNPDITIMVQSVWMDYRSTDRKTSPIWELNGSQRTKQYNHDY  
 SKAPDITTKAOGLEETAADLAAPGNIGLYSALADIGGTHKDISOVPNKYVKNPHH  
 GQVLDFTNKAETAKEPANKLGMWTQSFEDPHNSDRFFYGTGAGTGGTNTLNWDK  
 DTVSDIDFAESKPAATVRAKSISETTSWGVSTDAETWKEPAGTSGVKGPSTIV  
 STNGASIVFAETAGFTSTGNVFRFVRQGHLMAGVSGTSGWWRSTDDGKWT  
 KYSGVQEGDAVGFGRKSSGVPVFTSAKIDGVRGIFRSDDEGKTKRINDHNMWY  
 WTGASITGPDVGEVRVIGTNGRGIIVGDSSTPPPGQDSGASGNTGATPTPETPDS  
 DHNTGSDTDGTGETSPGTSGSTGAPDSGNTSGNTSSCKVRYSTTWDGSGFTGAATI  
 TNTSPSPNGTWLAFYPSGQTISVWSNATQTLSGRNVVLKNSGNWPTIPAGSVSEFG



sig\_peptide  
mat\_peptide  
FNGAYTSKNTPTPTSFSLNGVTALAD\*  
1..69  
70..2820  
/product="unnamed"

## ORIGIN

Alignment Scores:  
Pred. No.: 1,74e-57 Length: 2823  
Score: 2174.00 Matches: 430  
Percent Similarity: 59.48% Conservative: 141  
Best Local Similarity: 44.79% Mismatches: 311  
Query Match: 42.34% Indels: 78  
DB: 6 Gaps: 16

US-09-917-376-1 (1-957) x AM565635 (1-2823)

QY 10 ThrMetArgSerArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaValala 29  
DB 13 ACCGGTGGCCACACGACTAAGCTGCTGCTGCCACGACTGCGGTGGCATCGCGCC 72  
QY 30 AlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAlaHisAlaAlaThrThr 49  
DB 73 ACATCA-----ACCCCATGCGCAGTAAGTGCACCGCCCTCTTCTCATCACCT 123  
QY 50 GlnProTyrThrTyrSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleVal 69  
DB 124 TCCGGCTATAGCTGTGCGAACCTTGAGATTGGGCGCGGTTTGTGCTCGAATTGTC 183  
QY 70 PheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArg 89  
DB 184 TTTAACAGAACCCAGCCGCTTGGTGTATGTCGCGCAGATATTGGTGAGCGTACCGG 243  
QY 90 TrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnThr 109  
DB 244 CTCATGACTCCACCGGGGTTGGATCCTCTTACGGACCATATTGTTGGGATGACTGG 303  
QY 110 GlyTyrAsnGlyValSerIleAlaAlaAspProIleAsnThrAsnIysValTrpAla 129  
DB 304 AGCCACTCGGAATTCTGCTTGGCCACTGACCTGTTGACACCAACCGCGTGACCTT 363  
QY 130 AlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyValAlaIleLeuArgSer 149  
DB 364 GACAGAGGAGACCTCGGGAAGAGACGATGCTCGCGTTCGGTGGGAGGCAACATGCCG 423  
QY 150 AspGlnGlyAlaThrTrpGlnIleThrProLeuProPheIysLeuGlyGlyAsnMetPro 169  
DB 424 GACAAAGGCGAGACCTCGGGAAGAGACGATGCTCGCGTTCGGTGGGAGGCAACATGCCG 483  
QY 170 GlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPhe 189  
DB 484 GGTGCGGGATGGGAGACGCTCGCGCTTGCAGTATGCCCAACCAATAAAGTCCTGTACTTT 543  
QY 190 GlyAlaProSerGlyIysGlyLeuTyrArgSerThrAspSerGlyValAlaThrTrpSerGln 209  
DB 544 GGTGGGAATCGGGTAATGGACTATGGAATGCTCAACTGACTACGGAAACCTGGGGCAAA 603  
QY 210 MetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGly--- 228  
DB 604 GTGACGAGCTTCCCAATGCTGTAAGTACTAGTCGCC-----GATGCTTCTGTGGCG 654  
QY 229 TyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspIysSerSerSerLeu 248  
DB 655 TACACCGGCAAAACCAAGGTGTGTGGGTAACTTTTGACCCCATCATCAGCGCAAGCT 714  
QY 249 GlyGlnAlaSerIysThrIlePheValGlyValAlaAspProAsnProAlaPheTrp 268  
DB 715 GGTGACACACCAACCAACGATCTACGTAGGTGTAGCAGCAAGCAGAACACACGTTTATCGA 774  
QY 269 SerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyValaProThrGlyPheIlePro 288  
DB 775 TCTACCATGGCGGGGACATGCGACCGGCTCCCTGGGCAACCCACCGGATTCCTGGCG 834  
QY 289 HisIysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGly 308

DB 835 CAAAGGGAGTGTCTTGACCATAAAGGTCAGCAACTCTACATCGCTACTCTCAGATACTGGT 894  
QY 309 GlyProTyrAspGlySerSerGlyAspValTyrLysPheSerValThrSerGlyThrTrp 328  
DB 895 GGGCGGTATGACGGGTGGAAGGGGATGTGTGGCGCTAGATATTTCAGTGCCACGATGG 954  
QY 329 ThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGly 348  
DB 955 ACTCGAATCAGCCCAATTCCTCTACC---TCATCGAATTCAGCGTTCGGGTACTCCGGT 1011  
QY 349 LeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrp 368  
DB 1012 TTAGCCATTGACCGCAAAACCCGTGACACCATCATGTTGTTCCCAAGTGTCTGTGGTG 1071  
QY 369 ProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrpAsp 388  
DB 1072 CTGTGACATGATGATATATCGTCCACCGACCGGGGAAAACGTGTGCGCTATCTGGAA 1131  
QY 389 TrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrp 408  
DB 1132 CTGAACGGTTCAGCGCGGTACAAGACGATCAACCCACGATTACTCAGGGGCTCCCTGG 1191  
QY 409 LeuThrPheGlyValGlnProAsnProValProSerProIysLeuGlyTyrMetAsp 428  
DB 1192 CTTGACTTTGGCAATACAGCCAAAGAGCCTGAAGCTAAACCCCAAACTCGGTGATGACC 1251  
QY 429 GluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThr 448  
DB 1252 CAGTCTTCGAGATTGACCGGCATAACTCCGCGCTTCTTCTATGGCACCGGCGCGGC 1311  
QY 449 LeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaPro 468  
DB 1312 ATCTATGGCGGAACGAACCTCACAACTGGGATGAAGGCCAAAGTTGATATCACTGTC 1371  
QY 469 MetValIysGlyLeuGluThrAlaValAsnAspLeuIleSerProProSerGlyAla 488  
DB 1372 AAGCCCAAGGATCGAAGAAACCGCAGCCCAAGACCTTGCCTGCTCCCTCGAAACATT 1431  
QY 489 ProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaVal 508  
DB 1432 GGCCTATACTCAGCGCTGGCAGACATCGGTGGCTTCAACCCCAAGATATCTCTCAGGTG 1491  
QY 509 ProSerThr---IlePheThrSerProValPheThrThrGlyThrSerValAspTyrAla 527  
DB 1492 CCTAACAGTACTACTCAAGAACCCACACCATGACACCGTGACTCATTTGACTTCGCC 1551  
QY 528 GluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSerGlnProAsn 547  
DB 1552 GAAAGCAACCGCACTGCTGCTGCGTGCAGGCAAA-----TCTATTTCTGGGAAACC 1605  
QY 548 AspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluPro 567  
DB 1606 ACCTCATGGGTGGGGTTTCAACTGATGCTGGCGGAAAACGTGGAAAACCTGGCGCCACGCA 1665  
QY 568 GlyGlyValThrThrGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrp 587  
DB 1666 TCAGGGGTAAAGGCCCGCGCTCCATCATGTGTGTCAGTAAATGCTTATCCATTGTGTGG 1725  
QY 588 AlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAla 607  
DB 1726 GCCCG-----GAAGTGGCGCTCTCGAGTTCACCAACAGTGGCTCTCTCGGTGCA 1779  
QY 608 AlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProIysThr 627  
DB 1780 ACGGTGAGCGGCTTCCTCAATGTCAGGTGCGATCAGACCGAGTCAACGCTAATACT 1839  
QY 628 PheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGln 647  
DB 1840 CTTTACGGCTTCGTTGACGGAAAGTTCTATCACTCCCAAAATGGTGGGCTTCCTTCACG 1899  
QY 648 ProValAla---AlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaVal 666



```
Db 1900 GCCTCAGCTTTCACGGGGTTTCCACCACTGGA-----AACGTTGCTTTCCGTGCCGTG 1953
Qy 667 ProGlyLysGluGlyAspLeuThrLeuAla-----AlaSerSerGlyLeu 681
Db 1954 CCGGGCGCTCAAGGTCTATCTGCTTGTGGAGTATCGGAAGCACCTATGGGATG 2013
Qy 682 TyrHisSerThrAsnGlyGlySerSerThrSerAlaIleThrGlyValSerSerAlaVal 701
Db 2014 TGGCGGCTCACTGATGGCGGAAGACTGGACCAAGGTATCTGGAGTTCAGGAAGCGAT 2073
Qy 702 AsnValGlyPheGlyLysSerAlaProGlySerSerThrProAlaValPheValValGly 721
Db 2074 GCCGTGTTGTTCCGTAAGCAACCAAGTAGTAGCGGATACCCAGTGATTTTTCACATCCGCC 2133
Qy 722 ThrIleGlyGlyValThrGlyAlaThrArgSerAspAspCysGlyThrThrTrpValLeu 741
Db 2134 AAGATTGATGAGTTCGAGGAATTTCCGGTCTGACGATGAAGGCAAGACCTGGAACGC 2193
Qy 742 IleAsnAspAspGlnHisGlnThrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAla 761
Db 2194 ATCAACGACACCAACCAACCAAGTATGGACTGGCGCATCAATCACCGGTGACCCAGAT 2253
Qy 762 AsnLeuArgArgValThrIleGlyThrAsnGlyArgGlyIleValThrGlyAspIleGly 781
Db 2254 GTCGAAGGGCGCTACATTGTTGTTACCAACCGTTCGGCGCAATTATCGTTGGTGAT----- 2307
Qy 782 GlyAlaProSerGlySerProSerProSerProSerValSerProSerAlaSerProSerLeuSer 801
Db 2307 ----- 2307
Qy 802 ProSerProSerProSerSerProSerProSerProSerProSerSerProSer 821
Db 2308 -----TCTCCACTCCCGCCCGCCCGCTCAAGTAGCGGACGCGCATCG 2352
Qy 822 SerSerProSerProSerProSerProSerProSerProSerProSerProSerPro 840
Db 2353 GGAATACAGGGGCCACACCAAGCGCTGAACCCCTGCACCGATCAACAACACTGGCGGG 2412
Qy 841 SerAlaSerProSerProSerProSerProSerProSerProSerProSerProSerPro 860
Db 2413 TCCACTGATACGGGCGAACTTCAGAGCGCGGAACAGGAAGTGGATCTCGGAACCGCACCT 2472
Qy 861 SerProThrProSerSerSerProValSerGlyGlyValLysValGlnThrLysAsnAsn 880
Db 2473 GACTCGGGGAACACTCTCTCGGAATACCAAGTTCATGCAAGGTTCGGTATTCACACGACA 2532
Qy 881 AspSerAlaProGlyAspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySer 900
Db 2533 GACTGGGCGCAGGC-----TTCACTGGGCGCAGGCACCATCACTAAACAGTCCCCC 2583
Qy 901 SerSerValAspLeuSerThrValThrValArgTyr----- 912
Db 2584 AGCCCAATTAAACGGCTGACGCTGGCGTCTACTTACCCTTCGGGCGCAGCATCATGTTCC 2643
Qy 913 ---TrpPheThrArgAsp-----GlyGlySerSerThrLeuValThrAsnCysAspTrp 929
Db 2644 GTGTGGTCAGCAACTCAGACTCTGCTGGTCGCAATGTTGTCTCAAGAACTCTGTTGG 2703
Qy 930 AlaAlaIleGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 949
Db 2704 -----ATCCCAACGATACCT 2718

RESULT 6
LOCUS AE007608 12732 bp DNA linear BCT 01-JUN-2004
DEFINITION Clostridium acetobutylicum ATCC 824 section 96 of 356 of the
complete genome.
ACCESSION AE007608 AE001437
VERSION AE007608.1 GI:15023819
KEYWORDS
SOURCE Clostridium acetobutylicum ATCC 824
ORGANISM Clostridium acetobutylicum ATCC 824
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
```

```
Clostridium.
1 (bases 1 to 12732)
Nolling,J., Breton,G., Omelchenko,M.V., Markarova,K.S., Zeng,Q.,
Gibson,R., Lee,H.M., Dubois,J., Qiu,D., Hitti,J., Wolf,Y.I.,
Tatusov,R.L., Sabathe,F., Dounette-Stamm,L., Soucaille,P.,
Daly,M.J., Bennett,G.N., Koonin,E.V. and Smith,D.R.
Genome sequence and comparative analysis of the solvent-producing
bacterium Clostridium acetobutylicum
J. Bacteriol. 183 (16), 4823-4838 (2001)
21359325
MEDLINE
PUBMED
11466286
2 (bases 1 to 12732)
Childress,D., Zeng,Q. and Smith,D.R.
Direct Submission
Submitted (24-JUL-2001) GTC Sequencing Center Production,
Finishing, and Bioinformatics Teams, Genome Therapeutics Corp., 100
Beaver Street, Waltham, MA 02453-8443, USA
Location/Qualifiers
1. 12732
/organism="Clostridium acetobutylicum ATCC 824"
/mol_type="genomic DNA"
/strain="ATCC 824"
/db_xref="ATCC:824"
/db_xref="taxon:272562"
73. .2592
/gene="CAC0919"
73. .2592
/gene="CAC0919"
/codon_start=1
/transl_table=11
/product="Probably secreted sialidase; several ASP-boxes
and dockerin domain"
/protein_id="AAK78895.1"
/db_xref="GI:15023820"
/transl_table=11
/translation="MNKRIVSMVAGLSIIFTTGVTHISAAANKAAASVOSQYKWDN
AKIAGGVPAVIFNFKTLIYARTDMGAYRWDKANNKWIPIPTDFSDMTLGCES
IATPDIIDNVRVIAAGLYTNDWDENAVILSQDQKNTWKRYQLPFKVGNNPGRNG
ERLQIDPNDKILYLGBSGNGLKWSYDYGOTWSKVDNFPDGTGVQDPQNEYTADKV
GVWETFDPSFTGTSQTQMTVGAADTKGNIYTNDDGKTWSAVKGQPKGLPHHG
ILASDGMILYSYNTSGPYDGSQGVKINTKGWNTNITPPAVGDTKSGFGGISVDA
QKNTVVPDLVLRNMPDEEIRYSTDAGTKWPIWDWNGYPNRTGLYDGLTNDKKNVDISV
GKTGTVPPDLVLRNMPDEEIRYSTDAGTKWPIWDWNGYPNRTGLYDGLTNDKKNVDISV
KANGIECAVNDVVPTKQAQLLSAVGDDCGFYHDDITKVPKMMTTPNFSATTSIDY
AESVPNFVRVGNVDTSKNQDKDCGISYDCKMWFSAAGSNIISGVYKACTVAAGADAK
TIWSPREGAANAATDNGNKTWPCSGIPQGAQKASDRNPKFKFGLNGKPYISTDA
GATTQSSQTGLPTKGKGFPTVIGHEGDIWAGDKGLHSTDSGATFTKSGVSDAS
DVTGLGSKTDDGYPAIYMDATIDGTAGIFRSDDEGATWIRINDAHOYGSPPDYCTG
DPNKYRVFVGTNGRGIYVGDIDGSOPTTPSVTLGDVNDGDIINGRDVWVLTQYIAG
KDVTRDKAADVNGDGTVNGRDLMTLRQYISGKITSFSSSTK"
complement(2634. .2978)
/gene="CAC0920"
complement(2634. .2978)
/gene="CAC0920"
/codon_start=1
/transl_table=11
/product="Protein related to MIPH/DOPD protein family,
function in bacteria is unknown"
/protein_id="AAK78896.1"
/db_xref="GI:15023821"
/translation="MPFINSTVTVKLDKEKSTLVELKGIIELLPGKSEDWLMVGFK
DNYPLYPKGEKKDAAFIEVKIFGSAKSNKLYSAISSLMENELSIKDCIYITVD
EVSNNWNGSLF"
3104. .4252
/gene="CAC0921"
3104. .4252
/gene="CAC0921"
/codon_start=1
/transl_table=11
/product="Metal-dependent amidohydrolase"
/protein_id="AAK78897.1"
/db_xref="GI:15023822"
/translation="MTIEKOEFLKLLIGYRELLHHPHLSMKEVETTKRIKRWLYE
NDIKVLDPFIEVGVAEIEGEHEGKTVAIRADILDALPIEKTETLPSVSNYKGMHACG
```



gene	HDFTAAIIGTAIILSKRAEIIYGRVRIIFQPGEEETGKGSOYIIIEBGLVRDVCIFGM HNKPDIPVGTIGIRSGELMASVDREIKVIGSGGHAGIPNKCIDPIVASOIVASMTQ VYSRISIESVVISITRFNSGTTWNVSDSAEMEGTVRTFNNIRISIRNLMERTAK GIARALGKIFKMFPSVPLENDVRFDKLCEIAEAADNGYKYVEAKRNIGGEDFSFYQ TVVPSFFVMGVGDTGEWHKANYLNKEDALMAVANYFSDLALKVLSKK"									
	4594. .4863									
	/gene="CAC0922"									
	/db_xref="GI:15023823"									
CDS	4594. .4863									
	/gene="CAC0922"									
	/codon_start=1									
	/transl_table=11									
gene	/product="Uncharacterized small conserved protein, homolog of yfjA/yukE B subtilis"									
	/protein_id="AAK78898.1"									
	/db_xref="GI:15023823"									
	/translation="MPDKALECEDIKKSIEKSTTQLIDSWYGESCFAPEKTYLILDR KLNVTYILQYKALYDISETYQNTDGDIAKAMKAVNEASEQK"									
CDS	5056. .6348									
	/gene="CAC0923"									
	/gene="CAC0923"									
	/codon_start=1									
gene	/transl_table=11									
	/product="Predicted membrane protein; CF_1 family"									
	/protein_id="AAK78899.1"									
	/db_xref="GI:15023824"									
CDS	/translation="MQSLKVMRNGLLKINEAKOGLKSAVDKFSLKNSRSIEIEIKN LKRELKAVNLEGEIRKINSINIDYVGRFTDVKCKASKLTSGYTKRKVGLNFGFE EVERVPGIGVPIIDAVESVDRNKFVLTQTTITLTGAIGVCIVTLPEAGITWKA WDALSAAGSDNIGTAASAIYORKVERKSEEDSGYDLFQSIPIGGAGWYSAKILG VKNPSYWCSSGKTLYDRSATVTLNVGHSVTSLGKDVKEKLPAKKAESSANKTLEA ANNAKIHMDAMMATKINGINNAKKEVVDLKKLNKNGVKNTPFPYKQFYDKNLL KDGKNLPQLOQDLIIQDQLETFEKTNTLNKKNYKNSIRDLNKNYRIDKASLYNIPA GGTLNTEDVQGSFHFYPPGGEIKFIFEIKVVLGGR"									
	6345. .6923									
	/gene="CAC0924"									
	/db_xref="GI:15023825"									
CDS	6345. .6923									
	/gene="CAC0924"									
	/codon_start=1									
	/transl_table=11									
gene	/product="Predicted membrane protein"									
	/protein_id="AAK78900.1"									
	/db_xref="GI:15023825"									
	/translation="MIRFOKEMALPFLTISILIIICGDIEKIISFGIEFDLILCNLFFI GLGIIGFRIYRWYKQIKYKYTELLKEDRLDKIEYIIDIYRVGVVYNTAFYIDGVKC KLRFPOYCKNDNVEYYLEPIYKNGEFPKDFIYADKKNYKKSISTESTNEITECEBK AEKRRKRLNRYVNSRMELAKYBEEAKNK"									
CDS	7003. .8043									
	/gene="CAC0925"									
	7003. .8043									
	/gene="CAC0925"									
gene	/codon_start=1									
	/transl_table=11									
	/product="TPR-repeat-containing protein"									
	/protein_id="AAK78901.1"									
CDS	/db_xref="GI:15023826"									
	/translation="MENENVLRAQELNRLGLPIERNAGDEALKYFNKAIKEDENFKE AYLNKAALCLANNKYDDAVSCYDKLIIKYVNEGKAVFGKGNVWFFYKDKNIKKAILEYN RSIYLGEMSEGVYNLALCMESIGELDDAIKWLNRALIINDONIITLNKKAVALIKIK KYFALECEKRVNLKIEVDNEEVTHFAMLLGEMGKEKAPETISRGREALLGQMTLCY DKAVLFEKQKEFKALECTEKKALLFDGNNVLITLKKGDLLTYLKKIDEAKSTYDKILE FHNNMGEVFAKANLCMLLGEHEDAEFFKEIIERLSGSESVLINSYYYRNLNLKKWA KTRKQRLIRKQ"									
	7998. .8357									
	/gene="CAC0926"									
CDS	7998. .8357									
	/gene="CAC0926"									
	/codon_start=1									
	/transl_table=11									
gene	/product="TPR-repeat-containing protein"									
	/protein_id="AAK78902.1"									
	/db_xref="GI:15023827"									
	/translation="MGKNBEAKYKAEIKKNYFIILKYPYDVKLNFFKANCILRDIED									
gene	YKABELYEYIIDLNDALLEVYLMRARNRISLNKYEAAKEDLNTVTKLNPAYRKVIEL DQLKEFLSGKSESCVR"									
	8379. .8675									
	/gene="CAC0927"									
	8379. .8675									
CDS	/gene="CAC0927"									
	/codon_start=1									
	/transl_table=11									
	/product="Hypothetical protein, CF-2 family"									
gene	/protein_id="AAK78903.1"									
	/db_xref="GI:15023828"									
	/translation="MNSDEIRIDEKADKDDVIDKDKAVGEKVKETYNLIHLSNRHIEDG IKGSVADSVITTEYNKIYKVMKEMQDNMGNDVAAVNNIVSSFDAEDKMIKTKFYR"									
	8713. .9012									
CDS	/gene="CAC0928"									
	8713. .9012									
	/gene="CAC0928"									
	/codon_start=1									
gene	/transl_table=11									
	/product="Uncharacterized small conserved protein, homolog of yfjA/yukE B subtilis"									
	/protein_id="AAK78904.1"									
	/db_xref="GI:15023829"									
CDS	/translation="MVDRCDIEIDASIFEETIKYVSDSKVRLHDLCDLENELRMED TWEGDAKEFDSTFPFGYSAMKKDCMLBELIKELKLVKTSFESLDTMKNLDKK"									
	9712. .10335									
	/gene="CAC0929"									
	9712. .10335									
CDS	/gene="CAC0929"									
	/codon_start=1									
	/transl_table=11									
	/product="SAM-dependent methyltransferase"									
gene	/protein_id="AAK78905.1"									
	/db_xref="GI:15023830"									
	/translation="MGSMQYFDNIAASWNVIRSEYFEERLYKVLKSVISKIKITADF GCTGFISIALASDANNVRLKESAHDKLNNIYLLKSLTSLTSVTLFDE SIDAVFINMALHHIKDAEKAIKEMRYVLKDGTVI SDVSEHNGKWAREMFDEWLGF SNEQIDMWLKVGFKNIDIENTDLCKGTSRSGHTETSIFIACAAK"									
	10353. .11498									
CDS	/gene="CAC0930"									
	/note="metB"									
	10353. .11498									
	/gene="CAC0930"									
gene	/codon_start=1									
	/transl_table=11									
	Length: 12732									
	Matches: 392									
CDS	Percent Similarity: 64.39%									
	Best Local Similarity: 49.49%									
	Query Match: 39.67%									
	DB: 1									
US-09-917-376-1 (1-957) x AE007608 (1-12732)										
Qy	13 SerArgLeuValSerLeuLeuAlaA-----ThrAlaSerPheAla 27									
	76 AATAAAGGATTTGTTAGCATGTTGCAGGACTTAGCATATTTTCACAACTGGTTT--- 132									
Qy	28 ValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAlaHisAlaA 47									
	133 -----GTAACACATATAAGTCGACGACGATAATAAAGGACGAGCTTCA 174									
Qy	48 Thr---ThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAsp 66									
	175 GTTCAAAGCCCAAGGATATAGTGGGATAATGCAAGAGATAGGTGCTGCTGGATATGTTCCA 234									
Qy	67 GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGly 86									
	235 GCAGTTATTTAATAATAAACAAGAAAAGATTTATATATATATGCTCGTACACATATGGGAGGA 294									
Qy	87 MetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrp 106									



Db 295 GCATATAGATGGGCAAGCGTAATAAATGAAGTGAATTCGAATTAATCTGAT- - - - -GGAATT 348  
Qy 107 AenAsnTrpGlyTyrAsnGlyValValSerIleAlaAaAspProIleAsnThrAsnLys 126  
Db 349 TCTGATTCGCAAAATCGTTCGGCTGCGAAAGTATAGCAACTGACCCCAATAGACACTAATCGT 408  
Qy 127 ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 146  
Db 409 GTTATATAGCAGCAGGTTGTACACAATGATTCGCAAGATGAATAATGCATATATCTT 468  
Qy 147 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGly 166  
Db 469 AGTTCACAGGATAAAGGTAACACATGGAAGAGATACCAACTCCATTTAAAGGTAGAGGT 528  
Qy 167 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 186  
Db 529 AATATGCCGGGAAGAAATATGGGAGAAAGACTTCAANTAGATCCCTAATGATGATAAATA 588  
Qy 187 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThr 206  
Db 589 CTTTATCTTGAGCTAGAGTGGTAATGGACTTTGGAAAGTGAAGACTATGGACAACA 648  
Qy 207 TrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 226  
Db 649 TGGTCTAAGTTGATAATTCCTCGATACAGCGGACTATGTACAAGATCCTCAAAATGAA 708  
Qy 227 ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSer 246  
Db 709 -----TATACAGCAGATAAAGTTGGAGTAGTTTGGGAAACCTTTGATCCTTCAACAGGA 762  
Qy 247 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAsp- - -ProAsnAsnPro 265  
Db 763 ACAAAAGGAAGCCCAACTCAAACTATGATGTGTGGCAGCAGATAAGACAGAAACAAT 822  
Qy 266 ValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGly 285  
Db 823 ATTACGTTACTAATGATGGAGGAAGACATGCGAGTGAAGTAAAGGACAGCCAAAGGA 882  
Qy 286 PheIleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSer 305  
Db 883 TATCTTCACATCATGATGATATCTT- - -GCTTCGATGATGATGCTTTATATATCATATAGC 939  
Qy 306 AsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSer 325  
Db 940 AATACTTCGGACCATATGACGGCTCAGATGGACAAAGTTTGAAGTATAAATACCAAAACA 999  
Qy 326 GlyThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGly 345  
Db 1000 GGAGAAATGGACAAACATACTCACTCGAGTAGGTGACACTAAATCT- - - - -GGT 1050  
Qy 346 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 365  
Db 1051 TTTGGAGGAATATCAGTAGATGCACAAATCCAAATCAATTAATGTTGTAGTTGCAACTTAAAT 1110  
Qy 366 SerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArg 385  
Db 1111 AGATGGTGGCGGATGAAGAAATATACAGAAGTACAGATGCGAGGTAAGACCTGGAAACCT 1170  
Qy 386 IleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAla 405  
Db 1171 ATATGGGAATGGAAATGCCAATCCAAACCGTACTTTGGGATATACTTAGATTACTCAGCA 1230  
Qy 406 GluProTrpLeuThrPheGly- - -ValGlnProAsnProProValProSerProLysLeu 424  
Db 1231 CAACCTTGGCTTGTATGGGTGAAGACCGGGGTAAACACACACAGATCCATTTAGTTAAATTA 1290  
Qy 425 GlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGly 444  
Db 1291 GGCTGGATGATGGGAGACTTAGAAATTTGATCCATTTAACTCGGATAGAATGTTCTACGGT 1350  
Qy 445 ThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlnIle 464  
Db 445 ThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlnIle 464

Db 1351 ACAGGTGCAACTCTTTATGGAACTGATGATTTGACTAATTTGGGATAAAGTAAGAATGTA 1410  
Qy 465 HisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerPro 484  
Db 1411 GACATCTCTGTAAAGCAAAATGGAATGGAAGATGTGCGATAAATGATGTGTAGTACCA 1470  
Qy 485 ProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAsp 504  
Db 1471 ACAAAAGGCGCAACCTTTTAAAGCGCTGTTGGAGATGACTGCGGATTTTATCATGATGAC 1530  
Qy 505 ValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerVal 524  
Db 1531 ATAACCTAAAGTTCCAAAGTAAATGATGACACAACTTAACCTTTAGTGTCAACAACGACT 1590  
Qy 525 AspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSer 544  
Db 1591 GATTATGCAAAAGTGTCTTAACCTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1650  
Qy 545 GlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGln- - - 563  
Db 1651 AATCAACAGATAAAGATTTCTTACGATGGTGGTAAGAACTGTTTCTTCAGCT 1710  
Qy 564 GlySerGluProGlyGlyValThrGlyGlyThrValAlaAlaSerAlaAspGlySer 583  
Db 1711 GGAAGCAATATATCAGGAGTATTAAGAGCAGGAACCTGTAGCAGCTGGTGTGATGCTAAA 1770  
Qy 584 ArgPheValTrpAlaProGlyAspProGlyGlnProValValThrAlaValGlyPheGly 603  
Db 1771 ACTATAGTATGTTCTCCA- - -GAGGAAGGTGCTAATGTCAGCATATAGCAGATAATGGT 1827  
Qy 604 AsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgVal 623  
Db 1828 AATAAATGGACACCGCTGTTCTGGTCTCCACAGGAGCAAGGTAGATCAGACCGTGA 1887  
Qy 624 AsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGly 643  
Db 1888 AATCCAAAGAAATCTATGTTCTTAAATGGTAAATCTATATAGTATGATGTCAGGT 1947  
Qy 644 ValThrPhe- - -GlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMet 662  
Db 1948 GCAACATTTACTCAAAGTTTCAACACAGGTCTTCTCAAAAGGCAAA- - -GGAATC- - - 2001  
Qy 663 PheHisAlaValProGlyLysGlyAspLeuTrpLeuAla- - -AlaSerSerGlyLeu 681  
Db 2002 TTTAAGACTGTTATAGTTCATGAGGTGATATTTGATGCTGGAGGAAAGACCGATT 2061  
Qy 682 TyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaVal 701  
Db 2062 TGGCATTCACTGATCTCGTGCACATTCACAAAAGTATCTGTTGTTGATGCTTCTGAT 2121  
Qy 702 AsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValValGly 721  
Db 2122 ACAGTAGGTCTTGGAAATCCAAACACAGATGATGATGATGATGATGATGATGATGATGAT 2181  
Qy 722 ThrIleGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpValLeu 741  
Db 2182 ACATTTGATGGAACACCGGGAATATTCGGTTCAAGATGATGATGATGATGATGATGATGAT 2241  
Qy 742 IleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAla 761  
Db 2242 ATAATGATGATGCTCACCAATATGTTCCCGACACTACTGCACTAACTGGAGATCCTAAC 2301  
Qy 762 AsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGly 781  
Db 2302 AAATATGACGAGTCTTGTGGAACTTAATGGAAGAGGTATTTGTTTATGATGAGACATAGAT 2361  
Qy 782 GlyAlaProSerGlySerProSerProSerValSer 793  
Db 2362 GGTTCCT- - -CAACCTACACCTACACCAAGTTTACT 2394  
RESULT 7  
AJ585344  
LOCUS

AJ585344 2950 bp DNA linear BCT 11-OCT-2003







QY 404 SerAlaGluProTrpLeuThrPheGlyValGlnProAsnProProValProSerProlys 423  
 Db 1520 TCCGACACCGTGGCTGGATTGGGAACTGAGAAACAGCTGCCGGAATCAATCCGAAA 1579  
 QY 424 LeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyr 443  
 Db 1580 CTGGATGGATGATAGGTGACATAGAGATTGACCGGTTTAAATTCGACCGCATGATGAT 1639  
 QY 444 GlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGln 463  
 Db 1640 GTTACCGGTGCAACTATCTGGTGTGACACTTCTACTGCTGGACAGAGCGCGCAA 1699  
 QY 464 IleHisIleAlaProMetValIysGlyLeuGluThrAlaValAsnAspLeuIleSer 483  
 Db 1700 GTAAAAATCGAGTTAAAGCTACCGGAATAGAGAAATGTCGGTATTAGACCTGTTAAGC 1759  
 QY 484 ProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAla 503  
 Db 1760 CGCGCGAGGGTGCACCGCTTGAAGTGCAGTTGGCGACCTTGTCTGGTTTGTTCATGAT 1819  
 QY 504 AspValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSer 523  
 Db 1820 GACCTGAAAGTGGTCCGAAAAAATGACGTT---CCTTCTATTCTTCAGGTACGGA 1876  
 QY 524 ValAspTyrAlaGluLeuAsnPro-----SerIleIleValArgAlaGlySerPhe 540  
 Db 1877 ATTGATTATCGGACCTTGTCCGAACCTTATGGCATTTGTTGCAAGGCTGATTTGAT 1936  
 QY 541 AspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsn 560  
 Db 1937 GATGTA-----AAGAAGATTCTTCTCTATATGACGAGGAAGGAAT 1978  
 QY 561 TrpPheGln---GlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSer 579  
 Db 1979 TGGTTCACGCCACCTTAATGAGACACCAACCTGGTAGCGGGTTCGGTTCGCGTTGCA 2038  
 QY 580 AlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAla 599  
 Db 2039 GCGGATGCMAATCATGATTATTGGACACG---GAAATGCAAGTCTCGCAGTT---ACA 2092  
 QY 600 ValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArg 619  
 Db 2093 ACGGACACCGAAACTCATGGAAGTGTGTACAAATCTTGGTATGGTGGTGGTGGCA 2152  
 QY 620 SerAspArgValAsnProIysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSer 639  
 Db 2153 TCCGACCGGTGTAACCGTAAAAAATTTCTACGCAATTTCTATAACGCGCAAAATTTCTATAAAGC 2212  
 QY 640 ThrAspGlyGlyValThrPheGlnProVal---AlaAlaGlyLeuProSerSerGlyAla 658  
 Db 2213 ACGGACGGTGGATTAACTTTTACCATACAAGGACCGCAGCTTCCCAAGTCCGTTAAC 2272  
 QY 659 ValGlyValMetPheHisAlaValProGlyLysGlyAspLeuTrpLeuAlaAla--- 677  
 Db 2273 -----AAGATAAAAGCCGTACCGGCAAGGAAGGACATGATGGCTTCTGCTGCAAGA 2323  
 QY 678 SerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyVal 697  
 Db 2324 GAAGCGGATTTGGAGGTCCACTACCGTGGATATACCTTTTGAGAAAATCTCTCCAATGTT 2383  
 QY 698 SerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaVal 717  
 Db 2384 GACACGCTCATGTGTAGCTTCGGAAGGACGACCGGACAGGATTCATATGCGGATT 2443  
 QY 718 PheValValGlyThrIleGlyGlyValThrGlyAlaTyrArgSerAspAspCysGlyThr 737  
 Db 2444 TACATTACCGGTAATAATTGACAAATGTTTATGAGATTTCTTCGTTCCGATGATCCGCGCAG 2503  
 QY 738 ThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThr 757  
 Db 2504 ACATGGGTGCGTATCAACGACGACGACGATATGCGCTGTTGATATCGCAATAACA 2563  
 QY 758 GlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleValTyr 777

Db 2564 GGTGACCCGAGAGTATACGAGCTGTATATATTCGCACCAACGAGAGAGTATTGTTTAC 2623  
 QY 778 GlyAspIleGlyGlyAlaProSerGlySerProSerProSer 791  
 Db 2624 GCGGAA-----CCTGCTTCAGATGAGCCTGTACCCACT 2656  
 RESULT 8  
 ABI292929 Agaricus bisporus mRNA 2579 bp mRNA linear PLN 08-AUG-2000  
 DEFINITION Agaricus bisporus mRNA for CEL6 protein.  
 ACCESSION AJ292929  
 VERSION AJ292929.1 GI:9755896  
 KEYWORDS cel6 gene.  
 SOURCE Agaricus bisporus  
 ORGANISM Agaricus bisporus  
 Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;  
 Agaricales; Agaricaceae; Agaricus.  
 1  
 Morales-Almora,P. and Thurston,C.F.  
 Molecular analysis of the cellulolytic genes in Agaricus bisporus  
 Unpublished  
 2 (bases 1 to 2579)  
 Morales-Almora,P.  
 Direct Submission  
 Submitted (07-AUG-2000) Morales-Almora P., Microbiology, King's  
 College London, 150 Stamford Street, London, SE1 8WA, UNITED  
 KINGDOM  
 FEATURES  
 Location/Qualifiers  
 1..2579  
 /organism="Agaricus bisporus"  
 /mol\_type="mRNA"  
 /strain="D649"  
 /db\_xref="taxon:5341"  
 /clone="PMR152"  
 /dev\_stage="mycelium"  
 /country="United Kingdom"  
 1..2579  
 /gene="cel6"  
 28..2448  
 /gene="cel6"  
 /function="putative cellulase"  
 /codon\_start=1  
 /product="CEL6 protein"  
 /protein\_id="CA02964.1"  
 /db\_xref="GI:9755897"  
 /db\_xref="GOA:Q9P4T8"  
 /db\_xref="UniProt/TREMBL:Q9P4T8"  
 /translation="MPSRVKFWITLTGLTFLVYQAAQSQSVRWQNVKIGGGGFPVG  
 IVPNPEKGLAVARTDIGGAYKLNADDTWTPLLDPADDSRWNYWQVDALATDPVQPNR  
 LYLATGLTNSWDPSNGHLLISDYGKTFDAAPLPKLGSMFGRGMRGLVVDPNLN  
 SLIYFGARSGNLKWSKNSGRSKVTSFTDGTGSFVDPDSDTGLNSDKIGIAWTLID  
 KASGSGSATPRIFVGVANKGNSIYISNGSSWSAVAGQPTSFILPHKGVLSPSERA  
 LYITYSDGGPYDTSGAVYKYSIDTAWTDITPVSGSDLFFGFGGLALDTQKNVTVM  
 VAALNSWPDQIFRSTNGASWTPLTMDWAVYPLNKKYSYNAALAPWIGN1VDVTP  
 GNLOIWMWMEISIDPFDSNHWLYGTGETIYGRDILLKWDNAHNVTIKSLADGVEETS  
 VOALISPPSGPLVSAIGDVCGFVHQSLTRAPSAOFTNPKWTSADIDFAGNPSNVV  
 RIGDTSGKQVLAISDYGVTNWQHFQPDNVQGGKVAISADADILWRITNGNGVMV  
 SRNQATFNIFYGASGTFYVTDNNGKTFSAHSLGTSATSFVDITVHPVSGDIWASDT  
 KGLFHTSDSGTFSATIGTQAWGALGAPRSTGGYPAPVFAAANYGGEVAFPRSDRG  
 VNWVKINDAAHGFGAASANMAADPRVGRVYIGTNGRGI FYGDVAGSAPPPTSSATS  
 TTVSVSTTTTSTTTTSTTTTSSNGNSFVSGFGCGGCGTCTTCVTGTCVQVQNP  
 FYSQCIPI"  
 28..96  
 /gene="cel6"  
 97..2445  
 /gene="cel6"  
 /product="CEL6"  
 sig\_peptide  
 mat\_peptide  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 9e-46 Length: 2579  
 Score: 1784.00 Matches: 355



Percent Similarity:	59.78%	Conservative:	131
Best Local Similarity:	43.67%	Mismatches:	277
Query Match:	34.74%	Indels:	50
DB:	8	Gaps:	13
US-09-917-376-1 (1-957) x ABI292929 (1-2579)			
QY	20	LeuAlaAlaThrAlaSerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIle 39	
DB	28	ATGCCACGACGAGTTAAATCTCGGATAACTCTGACACACTTGGCACCTTCAGCCCTTCTCTAC 87	
QY	40	ThrAlaSerProAlaHisAlaAlaThrGlnProTyrThrTrpSerAsnValAlaIle 59	
DB	88	-----GGGCAAGCAGCTTCTCAGTCTTATAGTGGCAGAGATGTCAGATC 132	
QY	60	---GlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeu 78	
DB	133	GGCGGTGGTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 192	
QY	79	TyrValArgThrAspIleGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIle 98	
DB	193	TATGCTCGCACTGACATCGGAGGAGCTACAAAGCTGAAT--GCTGATGATCTTGGAGC 249	
QY	99	ProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAla 118	
DB	250	CCTCTACTTGATTTCTCGTGATGATTCCTGGTGGAAATTATGGGGGTGATGATGCTCTTGA 309	
QY	119	AlaAspProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrp 138	
DB	310	ACTGATCCGGTACACCCACCGCTTTTACCTAGCGACTGGGTGTGTATACCAACTCGTGG 369	
QY	139	AspProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThr 158	
DB	370	GACCAAGTAACGGCCATATCTCATATCAAGTGATATATGGTAAACATTCGATGCGCGG 429	
QY	159	ProLeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAla 178	
DB	430	CCTTTGCCATTCACGCTCGGTGGTAAATATGCGTGGGCGTGAATGGAGAGGCTCTCGTC 489	
QY	179	ValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrp 198	
DB	490	GTGATCCCAATCTGNAACAGCATCTTTATTTTGGCGCTCGCAGTGTAAAGGTCTCTGG 549	
QY	199	ArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThr 218	
DB	550	AAGTCAACAAATTCGGTGGAGCTGGAGCAAGGTAACAAAGCTTCACTGACACTGGGTCT 609	
QY	219	TyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrp 238	
DB	610	TTGTTCCCGATCCATCAGATGACACCGGCTGAAATTCGATAAGATTGGGATCGCATGG 669	
QY	239	ValAlaPheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGly 258	
DB	670	GTTCACATCGATAAAGCGTCAGGCTCTCTGGTTCGCTACTCTCCCGGTATATTTGTCGC 729	
QY	259	ValAlaAspPro---AsnAsnProValPheTrpSerArgAspGlyAlaThrTrpGln 277	
DB	730	GTTCGCAACAGGGCTCAACAGCATTTATATATCTACGAATGGCGGTCTCTCTGGAGT 789	
QY	278	AlaValProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsn 297	
DB	790	GGCGTGGAGGTCAACCAACTAGCTTCTCGCCACATAAAGGCGGTCTATCCCTCTGAG 849	
QY	298	HisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAsp 317	
DB	850	AGGGCACTTTATATCACTTACTCCGATGGTGGCGGTCTTACGATGGCAACAGCGTGCA 909	
QY	318	ValTrpLysSerValThrSerGlyThrTrpThrArgIleSerProValProSerThr 337	
DB	910	GTTTACAGTACAGCATTCAGCTGGCGCATGGCTGACATACACCGGTGACGGGAAGC 969	
QY	338	AspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsn 357	
DB	970	GATCTC-----TTCTTTGGTTTCGGAGGCTTAGCTTAGATACACAGAAAAACGGG 1020	
QY	358	ThrIleMetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThr 377	
DB	1021	ACTGTCATGGTTGCAGCCCTCACTCGTGGTGGCTGATGGTCAGATATTTCCGCTCGACC 1080	
QY	378	AspGlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeu 397	
DB	1081	AACGGGGTGCATCTTGGACTCGCTGTGGAGCTGGCAGTCTTACCGCAGCTCTCAACAAA 1140	
QY	398	ArgTyrValLeuAspIleSerAlaGluProTrpTrpLeuThrPheGlyValGlnProAsnPro 417	
DB	1141	TACTACAGTTATAACCTGCTCGCGCGCTGG-----ATAGAGCCCAACATC 1188	
QY	418	ProValProSerPro-----LysLeuGlyTrpMetAspGluAlaMetAlaIleAsp 434	
DB	1189	GTGTGATGTACCCCGGGAATCTACAAATCGGTGGATGATGATGATGATGATGATGAT 1248	
QY	435	ProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAsp 454	
DB	1249	CTTTTCGACTCGAATCATTTGGCTCTACGGCAGCTGGAGAACTATCTATGGAAGTCGTGAC 1308	
QY	455	LeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGlu 474	
DB	1309	CTCCTCAATGGGACAGTGCACAACTCACCATCAAGTCGCTCGCAGATGGTGTGCGAG 1368	
QY	475	GluThrAlaValAsnAspLeuIleSerProProSerGlyAlaProLeuIleSerAlaLeu 494	
DB	1369	GAACCTCGCTCAAGCCCTTATTTCACTCTTTCGGACCGCTCTTGTTCGCTATT 1428	
QY	495	GlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThr 514	
DB	1429	GGTGAGTTGGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1488	
QY	515	SerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIle 534	
DB	1489	AATCCCAAGTGGTGGACCTCGGTGACATCGATTTTCGCTGCAATAACCTTAGTAATGTC 1548	
QY	535	ValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSer 554	
DB	1549	GTTCGATTTGGTACTGGTGTAGTACTAGT-----GGCAACAGGTGCGCATCTCT 1599	
QY	555	ThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyGly 574	
DB	1600	TCCGACTATGGCGTTACTTGGAAATCAGCATTTGGAGCTCTGTATATGTTCAAGGCGGA 1659	
QY	575	ThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGln 594	
DB	1660	AAGTTTGGCATCTCCGCTGATGCGGACATCATCTCTGGCGTACAAATGGTAACGGT--- 1716	
QY	595	ProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAla 614	
DB	1716	----- 1716	
QY	615	AsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGly 634	
DB	1717	---GTCATGTGTGGGAAACCCAGGCCACCTTCAATATTTTCTACGGAGCCTCCGGCAAA 1773	
QY	635	ThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuPro 654	
DB	1774	ACATTTACGTTTCCACAGCAACGGCAAGACATTTTTCAGCCCATGGTAGCTAGGTCTCA 1833	
QY	655	SerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysGlyGlyAspLeuTrp 674	
DB	1834	GCTACGCTGTCTTCGACATAAATGCTCCAC-----CCGAGTGTACGGCGGTATCTCGG 1887	
QY	675	LeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerTrpSerAlaIle 694	
DB	1888	GCCTCCACTGCAAGGGCCCTATTCCATCTTACTGACAGCGGTGGCACTTTTAGTCTATC 1947	
QY	695	ThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyr 714	
DB	1948	TCGTGATACACAGGCTGGGTGTGCTCTTGGTGCACCTAGATCTACAGGAGGATAT 2007	







839 GCGAATGGCGTGGTCTCTTACGATGGCAGCAAAACGAAACAGTCCACAAAGTACCAACATCACC 898  
325 SerGlyThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPhe 344  
899 AGCGAGTGTGGCGGATATAAGTCCC-----ACCTCCCTTGGCGAGTACCTACTATAC 949  
345 GlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGln 364  
950 GGCACGGTGTCTGTCTCGGTGACCTCCAGATCTCTGGCACCTTGATGGTGGCTGCACCTC 1009  
365 IleSerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyValAlaThrTrpThr 384  
1010 AACTCTGGTGGCGGATGAGCTCATCTTCGTAGCACTGACTCTGGAGCGACCTGGTCC 1069  
385 ArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSer 404  
1070 CCTATTGGAGTGGACGGTATATCCGAGCATCAATTACTACTACAGCTATGATATCAGC 1129  
405 AlaGluProTrpLeuThrPheGlyValGlnProAsnPro---ProValProSerProLys 423  
1130 AACGCCCTTGGATCCAGGACAGCACTTCGCGCAACCAATTCCCGGTT-----CGC 1180  
424 LeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyr 443  
1181 GTCGGCTGGATGGTGAAGCCCTTGGCATTCACCAATTTGACTCCAACCACTGGCTCTAC 1240  
444 GlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrIleTyrAspSerGlyGlyGln 463  
1241 GGCACGGSCCTGACAGTATATGGCGGACACGACCTGACCAATTGGGACTCGAAGCACAA 1300  
464 IleHisIleAlaProMetValIysGlyLeuGluGluThrAlaValAsnAspLeuIleSer 483  
1301 GTGACGGTCAAGTCTCACTCGCGTCTGATTTAGGAGATGGCAGTCTCGACTGATCAG 1360  
484 ProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAla 503  
1361 CCGCCCGCGGACCGCGCTGCTGCTGCTGTGGCGACGCGCGGTTTACCACTCG 1420  
504 AspValThrAlaValProSerThrIlePheThrSerProValPheThrGlyThrSer 523  
1421 GATCTGGAGCGAGCGCCAAACCAACGCTACCAACGCGACCTACGCGACGACCAACGCG 1480  
524 ValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSer 543  
1481 ATCGACTACGCGGCAACAAGCTTCCCAACATCGTCCGTCGCGAGCTAGCGAC----- 1534  
544 SerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyAsnTrpPheGln 563  
1535 GACTATCCGACC-----CTGGCTCTGCTCGAATCTCGATCGACTTGGTATGCA 1585  
564 GlySerGluProGlyGlyValThrThrGlyThrValAlaAlaSerAlaAspGlySer 583  
1586 GACTAGCGCGCTCAACGACGACCGCGCGGCGGTGGCTCTTCCGCGCATGGTGCAC 1645  
584 ArgPheValTrpAlaProGlyAspProGlyGlnProValValThrAlaValGlyPheGly 603  
1646 ACGGTGCTTCTCATGTCACGACCTCGCGCGGTTGGTGTCAAGTCCGACGGC----- 1699  
604 AsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgVal 623  
1700 ---ACCTTGACTCGCGTTCCTCTCTCCGCTCCGGCGGCTCATCGCTCCGACAAAGTCC 1756  
624 AsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGly 643  
1757 GACAACAGGTCTTCTATGGTGGATCCCGCGGCGCAATCTAGTGTGGAAGAACACAGCG 1816  
644 ValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPhe 663  
1817 ACCTCTTCCACCAAGACCGTGAGC---CTGGGGTCCAGCACGACCGTCAACGCCATC--- 1870  
664 HisAlaValProGlyIysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHis 683  
1871 CGCGGCGCATCCCTCCATCCGCGGATGTCTGGCGGTGCGACGACCAAGGCGCTCTGGCAT 1930

684 SerThrAsnGlyGlySerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsn 702  
1931 TCCAGGATTACGGTAGTACGTTACGAGATCGGACGGCGTTCACGGCGGCTGGAGC 1990  
703 ValGlyPheGlySerAlaProGlySerSerTyrProAlaValPheValGlyThr 722  
1991 TTTGGTTTGGCAAGCGCTCTCCACGGCGCTACGTCGTGATCTTACGGGTTCTTCACC 2050  
723 IleGlyGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpValLeuIle 742  
2051 ATCGACGGTCCCGCGGCTGTTCAAGAGCGAGGACGAGGACCACTGGCAGGTGATC 2110  
743 AsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAla---IleThrGlyAspHis 760  
2111 TCGATCGCTCGACGAGATTCGGCTCG---GGTCCGCCACAGTGTGTGAACGGGACCTG 2167  
761 AlaAsnLeuArgArgVal-TyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle 780  
2168 CAGACCTACGGCGCGTGTTCGTGGGACGAAACGCGCGGCGCATCTTCTACGGCAAT-- 2225  
780 eGlyGlyAlaProSerGlySerProSerValSerProSerAlaSerProSerLe 800  
2226 -----CCAGCGGAGCTCTCGCGGCGCCACGCGCGGCGACGCGACGAC 2275  
800 userProSerProSerProSerProSerProSerProSerProSerProSerPr 820  
2276 ACCTCCAGACCTCG-ACGACGGTATCGACGAGTGTGAANAACCAACGACGAGCGCTC 2334  
820 oSerSerProSerProSerProSerProSerProSerProSerProSerProSerPr 840  
2335 GACCACTCTTCCAGCACCAACCGTGAAGACCACTAGTCTCGTCCAGCAGCTTCAA 2394  
840 oSerAlaSerProSerProSerProSerProSerProSerProSerProSerProSerPr 860  
2395 GCGCAGCTCGACCAACCAAGACTACCAACGACGACGACGACGACGACGACGAC 2454  
860 oSerProThrProSerProSerProValSerGlyGly 872  
2455 CACCGCGAGCGCTCGGCTACGCTCAGTGTGCGGC 2491

RESULT 10  
AY040839 3959 bp DNA linear PLN 01-APR-2002  
LOCUS Aspergillus niger endoglucanase C (eglC) gene, complete cds.  
DEFINITION AY040839  
ACCESSION AY040839.1 GI:19879405  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Aspergillus niger  
Aspergillus niger  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
REFERENCE 1 (bases 1 to 3959)  
AUTHORS Hasper,A.A., Dekkers,E., van Mil,M., van de Vondervoort,P.J. and de Graaff,L.H.  
TITLE EglC, a new endoglucanase from Aspergillus niger with major activity towards xyloglucan  
JOURNAL Appl. Environ. Microbiol. 68 (4), 1556-1560 (2002)  
MEDLINE 21914059  
PUBMED 11916668  
REFERENCE 2 (bases 1 to 3959)  
AUTHORS Hasper,A.A., Dekkers,E. and de Graaff,L.H.  
TITLE Direct Submission  
JOURNAL Submitted (18-JUN-2001) Molecular Genetics of Industrial Micro-organisms, Wageningen University, Dreijenlaan 2, Wageningen 6703HA, The Netherlands  
FEATURES  
source  
1. .3959  
/organism="Aspergillus niger"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:5061"  
<1038..>3841  
/gene="eglC"



```

mRNA      join(<1038..1137,1185..1329,1373..1571,1618..1858,
          1908..1961,2007.>.3841)
         /gene="eglC"
         /product="endoglucanase C"
CDS       join(1038..1137,1185..1329,1373..1571,1618..1858,
          1908..1961,2007..3841)
         /gene="eglC"
         /codon_start=1
         /product="endoglucanase C"
         /protein_id="AAK7727.1"
         /db_xref="GI:19879406"
         /translation="WRRTPAALVAGYLLDSVHAASQAYTKNNVTGCGGGTGPGLVF
NPSAKGVATRTDIGAYLRNSDDTWPLMDWANNNSNHWDGIDAIATDPVDTRRVY
AFYGSNGMLNDPDSILRETDQDGTWEETKLFPKWNNSPGRGVGERLAVDPDNLSI
YFGASNGNLKMTSDITGETWSNVTAFTKTCTYFQSSSTYTSDPVGIAWTFDTSGS
SGSPTRIPGVVVDTGYEVSDEAGETWTVSGEPMGFLPHKGILISPEHTLYIS
SNAGAPDVGCNTGVHKNYITSGVWTDDISPTSMSTDYVYGGLAVLDQVPGLVMVAALN
CWMPDELWRSTDSCGTWSPINAWNGXPISINYYSVIDSNAPWLDDTSTDDEFPRVG
WWVEALAIDFFDSHWHLYGTETIGYGHDLQNWDSHNVTIESLAVGISEMAVLGIT
PCPGKALIASVGDGDFYHTSLUTAPSYQHYHTPYSTSTNGIDYAGKNPKANVRSGSLS
SDPTLALSFFGESWADYAASSSTATQVALSADATTILLMNSDCGAYRSANSATLSA
VSSLPSGAVIASDKANNITYFYGASGSSFYLSDTATAFTVTTTLGSSTTANARAOPS
LAGDWVSDTCLGHFSTNYGKSPTOIGSCCTEGMSPGFKSPSSDGDPVPLRGFTVDG
VTGLFKTEQDSASNQWLIISDAERGFGFSASANVNGDLQNYGRVPFGTGRIPIYGDSPSG
TLPSTATASSASTSTSTSKVGSSTVSSSTAIIPTTSIKSLTIITLTITKS
SSSTTSTSSTATASAYQCQGGSGFTGTPQCPSGMTCTYENBYYSQCKSIPIGIATDR
G"

ORIGIN
Alignment Scores:
Pred. No.:           9,17e-43             Length:           3959
Score:              1690.00                Matches:            390
Percent Similarity: 53.46%                 Conservative:       128
Best Local Similarity: 40.25%               Mismatches:        315
Query Match:       32.91%                   Indels:            137
DB:                8                       Gaps:              26

US-09-917-376-1 (1-957) x AY040839 (1-3959)

Qy   18 SerLeuLeuAlaThrAlaSerPheAlaValAlaAlaAlaLeuGlyValLeuProIle 37
Db   1026 TCACTCGTACCACATCGCGGTACATTCGCGAGCCCTCGTAGCG--GGCTACCTG----- 1076

Qy   38 AlaIleThrAlaSerProAlaHisAlaAlaThrGlnProTyThrTrpSerAsnVal 57
Db   1077 -----CTGACTCGGTCCAACGCTGCCGCATCGCAGGCATATACCTGGAAGAATGTG 1127

Qy   58 AlaIle-----Gly 60
Db   1128 GTCAC-AGGGGGTATGACATATCTTGCATATATGAGGAGTTGACAAGCTAATTCATCAGGC 1186

Qy   61 GlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyVal 80
Db   1187 GCCCGAGGCTTCACGCCGGGTATCGTCTCAACCTTCAGCTAAGGTGCGCATATGCT 1246

Qy   81 ArgThrAspIleGlyMetTyArgTrpAspAlaAlaAsnGlyArgTrpIleProLeu 100
Db   1247 CGCACCCGACATCGCGGTGTCATATCGTTGAAC--TCGGACGACATACATGACCCCGTTG 1303

Qy   101 LeuAspTrpValGlyTrpAsnAsnTrp-----GlyTyArg-Asn----- 112
Db   1304 ATGGACTGGGCCAACCAACTCTAACTGTCGTAAAGCTTCTCTTACGAGATGCGCTAGGCTC 1363

Qy   113 -----GlyValValSerIleAlaAlaAspProIleAsnThrAsnLy 126
Db   1364 ACTGTCAGGCACGATTTGGGGCATATGATGCATCGGCACCGCCCGTCGATACCGATCG 1423

Qy   126 sValTrpAlaAlaValGlyMetTyThrAsnSerTrpAspProAsnAspGlyAlaIleLe 146
Db   1424 GGATATGTTGCAGTTGGCATGTACACCAACGACTGGGATCCCCAANTGACGGCTCCATCCT 1483

Qy   146 uArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGl 166

```



QY	476	ThrAlaValAsnAspLeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGly	495	QY	834	ProSerArgSerProSerProSerAlaSerProSerProSerProSerProSerProSer	853
DB	2594	ATGGCTGTTCGGGTCTCATCACTCCCGAGCGGCGCTCTCTCTCTCGGTGCGC	2653	DB	3626	ATTAAAGTCACCACTTTGACGACGACTACAAAGTCGAGTAGTTCGACTACTTCGACGCT	3685
QY	496	AspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSer	515	QY	854	SerSerProSerSerSerProSer	863
DB	2654	GACGATGGGGCTTCTACCACTTCCCTGACCGCGCCCATCAATACTACTACCAACC	2713	DB	3686	TCCACTGTCTACTGGGACGCGCTCGCATATGGCGAGTCGGTGGAGCGGGTTACGGGA	3745
QY	516	ProValPheThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleVal	535	QY	864	ProSerSerSerProValSerGlyGlyValLyValGlnTyrLyAsnAsn	880
DB	2714	CCGACTTACAGTTCAACAAACGCGTACTACGCGCGCAACAGCGCGCTAACATCGTC	2773	DB	3746	CCGACGAGTGGCGG-----AGTGGTGGACGTCGACGATGAGATGATATTATTCG	3799
QY	536	ArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThr	555	QY	881	-----AspSerAlaProGly	885
DB	2774	CGCTCTGATCCAGCGAT-----TCGACCCCTACTCTGCTTTATCTCTCC	2818	DB	3800	CAGTGTAAAGTCTATCCCTGCT	3820
QY	556	AspGlyGlyLyAsnTyrPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThr	575	RESULT 11			
DB	2819	AGTTTCGGCGAATCTTGTGACCGCACTAGCTGCTTCTTCTTCTACTCGGACGGCCAA	2878	AKI10506			
QY	576	ValAlaAlaSerAlaAspGlySerArgPheValTyrAlaProGlyAspProGlyGlnPro	595	LOCUS			
DB	2879	GTGCGCTCTACGCGATCGACACCATCTCTCTCATCAACAGCGACGCT-----	2929	DEFINITION			
QY	596	ValValTyrAlaValGlyPheGlyAsnSerTyrAlaAlaSerGlnGlyValProAlaAsn	615	ACCESSION			
DB	2930	---GCTATCTGTTCCGCAACCTCGCTACTCTTAGCGCACTCTCTCTCTGCGCTCGA	2986	VERSION			
QY	616	AlaGlnIleArgSerAspArgValAsnProLyThrPheTyrAlaLeuSerAsnGlyThr	635	KEYWORDS			
DB	2987	GCTGTATCGCTCCGCAACAGCTTACACACTTCTTCTGAGGCTCTCGGCTCTTTC	3046	SOURCE			
QY	636	PheTyrArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSer	655	ORGANISM			
DB	3047	TCTATCTTCTATCCGACACCGCGCGACTTC---ACTGTACACCAACCCCTCGCGACT	3103	REFERENCE			
QY	656	SerGlyAlaValGlyValMetPheHisAlaValProGlyGlyGluGlyAspLeuTyrLeu	675	AUTHORS			
DB	3104	AGCACAAACCGCAACGCCATC---CGCGCGCAACCTCTCTCGCAGCGATGTCGGTCT	3160				
QY	676	AlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIle---	694				
DB	3161	AGTACGACACAGAGTCTTCTCCACTCAACACTACGCGCAATCTTCCACCCAGATGGG	3220				
QY	695	ThrGlyValSerSerAlaValAsnValGlyPheGlyLeuSerAlaProGlySerSerTyr	714				
DB	3221	AGCGGATGCACAGCGGTGGAGTTCGGATTCGGAAGCCCTCTCGGACGCGACTAC	3280	TITLE			
QY	715	ProAlaValPheValGlyThrIleGlyGlyValThrGlyAlaTyrArgSerAspAsp	734	JOURNAL			
DB	3281	CGGTCTATTGGATTCTTTACCGTCGATGGTGTACCGGTCTCTTCAAAACCGAAGAT	3340	MEDLINE			
QY	735	CysGlyThrThrTyrValLeuIleAsnAspGlnHisGlnTyrGlyAsn---TyrGly	753	PUBMED			
DB	3341	CAGGGGTAAATTTGGCAGATAATTTCCGACCGCGACGCGGTTCGGGTCTGTCGGCA	3400	REFERENCE			
QY	754	GlnAlaIleThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArg	773	AUTHORS			
DB	3401	AATGTAGTCAATGGGATCTTCAGACTATGCGAGGGGTGTGTGTTGGCAATGGCGGA	3460				
QY	774	GlyIleValTyrGlyAspIleGlyAlaProSerGlySerProSerProSerValSer	793				
DB	3461	GGATCTCTTATGTTGAT-----CCGAGTGGAACT---CTGCGGAGTGGCACT	3505				
QY	794	ProSerAlaSerProSerLeuSerProSerProSerProSerProSerProSerProSer	813				
DB	3506	GCTACTCGCTAGTGTCTGTCGACAGCGGTGAAGAGTTTCGACCTCGACTTCGACATCC	3565				
QY	814	ProSerProSerSerProSerProSerProSerProSerProSerProSerProSer	833				
DB	3566	AAGGTAGTTCCAGCACTGTTTCTTTCGAGTACCGCTACAACTACTACTACGAGCAGT	3625				



Yoshimura, A.  
 Direct Submission  
 Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of  
 Agricultural Sciences, Department of Molecular Genetics, Head of  
 Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki  
 305-8602, Japan (E-mail: shikuchienias.affrc.go.jp,  
 Tel:81-29-838-7007, Fax:81-29-838-7007)  
 This clone is one of the 29K full-length cDNA clones from japonica  
 rice.  
 URL : http://cdna01.dna.affrc.go.jp/cDNA/  
 NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K.,  
 Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J.,  
 Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T.,  
 Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T. and  
 Yamamoto, M.

FAIS Genome Sequencing & Analysis Group: Ootomo, Y., Iida, Y.,  
 Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M.,  
 Kodama, T., Kuroaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J.,  
 Mizuno, K., Narikawa, R., Niikura, J., Oka, M., Ryu, R., Sugano, S.,  
 Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S.,  
 Yoshimura, A., Matsubara, K. and Murakami, K.  
 Genome Exploration Research Group in Riken Genomic Sciences Center  
 and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K.,  
 Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T.,  
 Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K.,  
 Hiraoka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y.,  
 Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J.,  
 Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,  
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
 Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N.,  
 Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H.,  
 Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,  
 Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F.,  
 Takaku-Akahira, S., Tanaka, T., Tomaru, A., Tota, T., Waki, K.,  
 Yasunishi, A. and Hayashizaki, Y.

## FEATURES

Location/Qualifiers  
 1..2724  
 /organism="Oryza sativa (japonica cultivar-group)"

/mol\_type="mRNA"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /clone="002-167-F04"

## ORIGIN

## Alignment Scores:

Pred. No.: 3.36e-41 Length: 2724  
 Score: 1633.00 Matches: 365  
 Percent Similarity: 58.11% Conservative: 151  
 Best Local Similarity: 41.10% Mismatches: 307  
 Query Match: 31.80% Indels: 65  
 DB: 8 Gaps: 27

US-09-917-376-1 (1-957) x AK110506 (1-2724)

QY 37 IleAlaIleThrAlaSerProAlaHisAlaAlaThrThrGlnProTyrThrTrpSerAsn 56  
 DB 88 ATCTCGATAACCGCTCTCCG-----GCCGTGGTCGAGCGGCTCTACTCTGGAAGAAT 141  
 QY 57 ValAlaIle---GlyGlyGlyGlyPheValAlaPheGlyIleValPheAsnGluGlyAlaPro 75  
 DB 142 GCCAAGTTGGAGCGCGGTGGCTTTGTGCGCGGCATTTGATTTCATCTACCCAGGCT 201  
 QY 76 GlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaAsnGly 95  
 DB 202 GCGGTGTCTTATGCTCGAAGCATATAGAGGCGCTGTACAGTGGAAAC---AGTGATGAC 258  
 QY 96 ArgTrpIleProLeuLeuAspTrpVal-----GlyTrpAsnAsnTrpGlyTyr 111  
 DB 259 TCGTGGAGCCCATCACCAGCTCCATCCAGCAATGCGAATGCGCAAAATTGG----- 312  
 QY 112 AenGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 131  
 DB 313 ---GGAGTGGATGCGCTTCTTGTATCCCGAGGACCCCAATGTGTGTACTTCGCCGGG 369

QY 132 GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGln 151  
 DB 370 GGTGTGTACCACTCTCTGGATCTAGCCCGCGGCATCTACAAGAGCAGGACAAG 429  
 QY 152 GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArg 171  
 DB 430 GGTGCGACTTGACATCTTCAACGCTCAGCTTCAAGGTGCGTGTAAACATGCTGCGCGC 489  
 QY 172 GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 191  
 DB 490 GGCATGGCGGAGCGCTGCTGCATCCCAAGAACTCCAACATTTGTATTTCGCGCGCA 549  
 QY 192 ProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 211  
 DB 550 CGCTCTGGAACCGGCTCTGGAGAGACACCGAGCGCGGCAACACTTCAAGAGGTGCGAC 609  
 QY 212 AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrTrpGlyTyrGlnSer 231  
 DB 610 TCCTTCAAGGCTGTGGTACCTTACATCCGCGTCCCGAGCGCAGCAGCGGCTTACAACAGC 669  
 QY 232 AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGlnAla 251  
 DB 670 GACATCAACGCGCTGACCTTTGTGACCTTTGACTCGCCTCGGCAACCATCAATGTTGCC 729  
 QY 252 SerLysThrIlePheValGlyValAlaAsp---ProAsnAsnProValPheTrpSerArg 270  
 DB 730 ACTGCGCGCATATTCGTAGGACCCCGCATACGTCACGGCTCGCTTACGTTTCTGAG 789  
 QY 271 AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHisLys 290  
 DB 790 GATGCGGCGCAGCAGATGGAACGCGGTCAAGGGCCAGCGCTTCTTCTCATTAAG 849  
 QY 291 GlyValPheAppProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyPro 310  
 DB 850 TGCAGGCTGCGAGCCCGAGGCGCACTCTTACGCGCGAGCAACGGGCGCT 909  
 QY 311 TyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpTrpArg 330  
 DB 910 TACGATGTGGCTCTGGCGCTCTGTGGCGATACGGCTCGCGCAGCGGACGAGCGGAC 969  
 QY 331 IleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThr 350  
 DB 970 ATCACTCTGCTCCGTTAGCGACTGACC-----TACGGATTGCGCGGCTCTCC 1020  
 QY 351 IleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAsp 370  
 DB 1021 GTGGACATGAAGAAGCCAGAACTCTGGTCTGTCGCGAGTCTCAACTCTTTGGTGGCTGAT 1080  
 QY 371 ThrIleIlePheArgSerThrAspGlyAlaThrTrpThrArgIleTrpAspTrpThr 390  
 DB 1081 GCGCAGATCTTCGCGAGCACCGACTCTGCGAAGAGCGTGTCAAGGCTCTGGCAGTGGGT 1140  
 QY 391 SerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThr 410  
 DB 1141 AACTATCCGAGATGATCTACTATTACGGCATCTCGACTCCCAAGGCCCTCTGG----- 1194  
 QY 411 PheGlyValGlnProAsn---ProProValProSerProLysLeuGlyTrpMetAspGlu 429  
 DB 1195 -----ATCCGCAAGAACTTCTCTCGATACCGACTCCAAGTGGCTGGGCTGATGATCGAG 1248  
 QY 430 AlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeu 449  
 DB 1249 TCTTTCGAGATCAACCGGACTCGACGCGCACTTCTCTGTACGGCACCGGACTGACCCCTC 1308  
 QY 450 TyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMet 469  
 DB 1309 TACGTTGGCCACGACTGACCACTGGGACCAACCAAGCGCAACATCACTCCAGTCCCTG 1368  
 QY 470 ValLysGlyLeuGluThrAlaValAlaAsnAspLeuIleSerProProSerGlyAlaPro 489  
 DB 1369 GCCGACGCGATCGAGGAGACCGCTGTTCAGGAGGTGGCTCGGCCCGCGCGCAGCGAG 1428



```
QY 490 LeuileSerAlaLeuGlyAspLeuGlyGlyPheThrHis-----AlaaspValThr 506
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1429 CTAATGGCTGCACGGCGGATATCCCGCTTCACTTCGCATCGCGCGACACCTTGA 1488
QY 507 AlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyr 526
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1489 ACTTCACCAAGCGGTCTGGGACAAACCCCAAGTTTCAACACCTCAGCGAGGTGACTAC 1548
QY 527 AlaGluLeuAnProSerIleIleValArgAlaGlySerPheAspProSerSerGlnPro 546
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1549 GCGCGCAACTCGCCGGAAGGTTGTGCGGTGGGCAAC---GATCGCGCGCTCAG--- 1602
QY 547 AsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGlu 566
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1603 -----ATGTCGCGCAGCTCGACGATGCGCGCTCACATGG-----AAGTGTATC 1647
QY 567 ProGlyGlyValThr-----ThrGlyGlyThrValAlaAlaSerAlaAspGlySerArg 584
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1648 CCGCGCGCGACACGACGACGCGGCGGTGCGGTCTCTACTCTCGGATGTCACACG 1707
QY 585 PheValTrpAlaProGlyAspProGlyGlnProValValThrAlaValGlyPheGlyAsn 604
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1708 ATGTTGTGTCACCGCGCACCGGTGCG-----GTGTACCGCTCCGTGAACACGAGC 1758
QY 605 SerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsn 624
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1759 AGTTTCTCGCGCTCTCTCGTTGCAGAGCGGAGTGTCTGCGCTCGGACGACGCAAC 1818
QY 625 ProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyVal 644
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1819 ACAAGTACTTCTACGGCGCATCCGGAGCACTTCTAGCTCTCGCTGACGCGCGCGTG 1878
QY 645 ThrPheGlnProValAlaAlaGlyLeuProSerSerGlyValAlaValMet---Phe 663
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1879 ACCTTT-----GCGCGCGCGCTCGCTCAGCGCGCCACGTCGATCCGCGACATC 1929
QY 664 HisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHis 683
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1930 GCGCGCCACCGCACGCGCTGCAAGCTTTATGTGTGTCGACCAACGACGCGCTCTTACC 1989
QY 684 SerThrAsnGlyGlySerSerTrpSerAlaIleThr---GlyValSerSerAlaValAsn 702
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1990 AGCAGCGATTTGGCGAGACCTTTACCAAGACTTCGCGCGCTGAGCGACACGATCCAG 2049
QY 703 ValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValValGlyThr 722
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2050 ATCGCGTGGCTGGCTCT---GGTTCCAAGTGG---AAGTGTACGCTTTGGCAGC 2103
QY 723 IleGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpValLeulle 742
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2104 GCGCGCAACGCGCGAGGTGTACGCGACGCGCGGCGGTCTGCTCGACTGATCTC 2163
QY 743 AsnAspAspGlnHisGlnTyrGly-----AsnTrpGlyGlnAlaIleThrGlyAspHis 760
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2164 CAGGCGCAGCAGCAGCGCTTCGCGATCATCAAC---GGATGAGCGCTCGCTGCGAGTGG 2220
QY 761 AlaAsnLeuArgArgValTyrIleGlyThrAsnGlyValGlyIleValTyrGlyAsp--- 779
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2221 AAGTCGCCAACAGGTATACGTGCGGACCAACGCGCGCGGTCTTCCAGGCGCAGCT 2280
QY 780 ---IleGlyGlyAlaProSerGlySerProSerProSerProSerProSerProSerPro 798
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2281 TCCATCTCCGAGGCGCGCGCAGCAGCCGCTCTGTGTACACACCACTCCGCGCGCAGC 2340
QY 799 SerLeuSerProSerProSerProSerProSerProSerProSerProSerProSerPro 818
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2341 TCGGCGCGCGCACCGACGACGAGGATGATCGCTAGCAGCTCCGCTTCGACTTCC 2400
QY 819 SerProSerSerSerProSerProSerProSerProSerProSerProSerProSerPro 838
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2401 ACCCGAGCTTCAAGCTGTGACTGCGACGACCACTTCCAGAGTCTCCTCGACCCCTG 2460
QY 839 SerProSerAlaSerProSerProSerProSerProSerProSerProSerProSerSer 858
```

```
Db 2461 GTCACCTTCGCTCTCTCCACCAAGACG-----TCCTTCGCGCGCGTGGGA 2505
QY 859 SerProSer-Pro-----ThrProSerSerProValSerGlyGlyValGlyValG 876
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2506 GGACGACGCGCGGAACTTACGAGCAGTGCAGCGGACGAGCGCTGGACAGCGCGC 2563
QY 876 nTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysProGlyLeuGlnValVa 896
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2564 -----AGTGTCTCTCCGATGACGCTGCAAGAAATGGAATG----- 2599
QY 896 lAsnThrGlySerSerSerVal 903
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2600 ----ATTGTACTCTCCCAAGTGTG 2617

RESULT 12
AY281371 2517 bp mRNA linear PLN 15-JUN-2003
Hypocrea jecorina strain QM6a Cel74a (cel74a) mRNA, complete cds.
ACCESSION AY281371
VERSION AY281371.1 GI:31747159
KEYWORDS
SOURCE
ORGANISM
Hypocrea jecorina (anamorph: Trichoderma reesei)
Hypocrea jecorina
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
REFERENCE
1 (bases 1 to 2517)
Foreman, P.K., Brown, D., Dankmeyer, L., Dean, R., Diener, S.,
Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., England, G.J.,
Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchinson, C.,
Olivares, H.A., Teunissen, P.J.M., Yao, J., and Ward, M.
Transcriptional Regulation of Biomass-Degrading Enzymes in the
Filamentous Fungus Trichoderma reesei
J. Biol. Chem. (2003) In press
2 (bases 1 to 2517)
Foreman, P.K., Brown, D., Dankmeyer, L., Dean, R., Diener, S.,
Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., Mitchell, T.,
Olivares, H.A., Teunissen, P.J.M., Yao, J. and Ward, M.
Direct Submission
Submitted (22-APR-2003) Genencor Intl., 925 Page Mill Road, Palo
Alto, CA 94304, USA
FEATURES
Location/Qualifiers
1..2517
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM6a"
/db_xref="taxon:51453"
/notes="anamorph: Trichoderma reesei"
1..2517
/genes="cel74a"
/genes="cel74a"
/notes="putative secreted endoglucanase"
/codon_start=1
/product="Cel74a"
/protein_id="AAE57752.1"
/db_xref="GI:31747160"
/translacion="MKVSRVALVLGAVIPAHAAFSWKNVKLGGGGFVPGIIFHPKPT
KGVAARTDIGLYRLNADDSWTAVDGIADNAGWNWIGIDAVALDPDDQKVAAVG
MTNSWDPSNGAIIRSSDRGATWSFTNLPPKVGNGMPGAGERLAVDPANSNIYFG
ATKRGWLKSTDGVTFSKVSFTATCTYIPDPSDSNGYNSDKGLMWVTFSTSTT
GATSRI FVTADNITASVVTNAGTSWNAVPGQPKYPPHKKLQPAEKALYLTYS
WDPADQLFRSTDSGTTWSWASYPTETYSISTPKAPKINKNIFDVTFSESPDG
LIRKLGWMTESLIDPDSNHMLYGTMTIFGGHDLTNWOTRHVSIQSLADGIEEFS
VODLAPAGSSELLAAGDDNGFTFASRDLGTSPTWATPTWATSTVVDYAGNSVK
SVVRGNTAGTQQVAISSDGGATWSIDYADTSMNGGTVAVSADGDTILWSTPGLGSAG
RSQFQSFASVSLPAGAVTASDKTNSVFYAGSGTSFYVSKDTSFTRPGLGSAG
TRIDIAAHPTAGTLVSTDVGI FRSTDSTGTFQGVSTALNTIQTALGVGSGSNWNL
YAFGTGPGSARLYAGSDSGSWTDIQSGQGSIDSTKVGSGTAGQVYGTNGRGV
FYAQGTVGGTGGTSSSTKQSSSTSSASSSTLRSVSVSTTRASTVTSRTSAAAGP
TSGSVAGHYAQCAGIGTGWGTQCVAPVVCQKQNDYYQCV"
gene
CDS
ORIGIN
```







QY 689 SerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsnValGlyPheGlyLys 707  
Db 1963 AGCAGCTTTGGCAAGTCTCCACGCCCTGACCAACACCTACCATGCCCTGGGTGG 2022  
QY 708 SerAlaProGlySerSerTyProAlaValPheValValGlyThrIleGlyGlyValThr 727  
Db 2023 GGCTCA---GGCTCAAGTGG---AAGCTGATGCTTCGGACCC-----GGCCCGTCA 2070  
QY 728 GlyAla-----TyArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAspAep 745  
Db 2071 GGGGCTCGCTCTACGCCAGTGGACAGCGCGCTCTCTGACGACATCCAGCGCTCC 2130  
QY 746 GlnHisGlnTyArgAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArg 765  
Db 2131 CAGGGCTTCGGCTCATCGACAGCACCAAGTTCGCGCGCAGCGACACCGCCGGCAA 2190  
QY 766 ValTyrlleGlyThrAsnGlyVargGlyIleValTyR-----GlyAspIleGlyGlyAla 783  
Db 2191 GTTACGTGGGGCACCACAGCGCGGGCGTCTTTTACGCTCAGGGAACCGTCGCGGGCGC 2250  
QY 784 ProSerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerProSer 803  
Db 2251 ACGGGCGGACr-----TCTCTCGTCCAGCAAGCAGACG 2283  
QY 804 ProSerProSerSerProSerProSerProSerProSerProSerSerProSerSerSer 823  
Db 2284 AGCAGCAGTACCTCTTCGCCAGCTCGACACACCGCTGAGCTGAGCGTGTATCCAG 2343  
QY 824 ProSerProSerProSerProSerProSerProSerProSerArgSerProSerProSerAlaSe 843  
Db 2344 ACCCGGGCTTCGACGGTGTCTTCGAGGACGCTCGCGCGCGTCCCGGCGGTCA 2403  
QY 843 rProSerPro 846  
Db 2404 GGGGTGCGCG 2413  
RESULT 13  
AP005031/c  
LOCUS  
DEFINITION 299550 bp DNA linear BCT 10-MAY-2003  
Streptomyces avermitilis genomic DNA, complete genome, section 11/30.  
ACCESSION AP005031 BA000030  
VERSION AP005031.1 GI:29606108  
KEYWORDS  
SOURCE Streptomyces avermitilis MA-4680  
ORGANISM Streptomyces avermitilis MA-4680  
REFERENCE 1  
AUTHORS Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C., Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T., Kikuchi,H., Shiba,T., Sakaki,Y. and Hattori,M.  
TITLE Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)  
MEDLINE 21477403  
PUBMED 11572948  
REFERENCE 2  
AUTHORS Ikeda,H., Ishikawa,J., Hanamoto,A., Shinose,M., Kikuchi,H., Shiba,T., Sakaki,Y., Hattori,M. and Omura,S.  
TITLE Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis  
JOURNAL Nat. Biotechnol. 21 (5), 526-531 (2003)  
MEDLINE 22608306  
PUBMED 12692562  
REFERENCE 3 (bases 1 to 299550)  
AUTHORS Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C., Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T., Kushida,N., Director-General of Biotechnology Center, Shiba,T., Sakaki,Y. and Hattori,M.  
TITLE Direct Submission

## JOURNAL

## COMMENT

Submitted (29-MAR-2002) Director-General of Biotechnology Center, National Institute of Technology and Evaluation, Biotechnology Center: 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan (E-mail: bio@nitech.go.jp, URL: http://www.bio.nitech.go.jp/, Tel:81-3-3481-1933, Fax:81-3-3481-8424)  
This work was done in collaboration with Haruo Ikeda (\*1), Jun Ishikawa (\*2), Akiharu Hanamoto (\*3), Chigusa Takahashi (\*3), Mayumi Shinose (\*3), Hiroshi Horikawa (\*4), Hidekazu Nakazawa (\*4), Tomomi Osonoe (\*4), Norihiro Kushida (\*4), Hisashi Kikuchi (\*4), Tadayoshi Shiba (\*5), Yoshiyuki Sakaki (\*6, \*7), Masahira Hattori (\*1, \*7) and Satoshi Omura (\*1, \*3).  
Final finishing process and all annotation were done by H. Ikeda and J. Ishikawa.  
\*1 Kitasato Institute for Life Sciences, Kitasato University  
\*2 National Institute of Infectious Diseases  
\*3 The Kitasato Institute  
\*4 National Institute of Technology and Evaluation  
\*5 School of Science, Kitasato University  
\*6 Institute of Medical Science, University of Tokyo  
\*7 RIKEN Genomic Sciences Center  
Following url is also available.  
http://avermitilis.is.kitasato-u.ac.jp.

## FEATURES

## source

1. .299550  
Location/Qualifiers  
/organism="Streptomyces avermitilis MA-4680"  
/mol\_type="genomic DNA"  
/strain="MA-4680"  
/db\_xref="taxon:227982"  
/notes="This strain is also named as strain: ATCC 31267, NCIMB 12804 or NRRL 8165.--synonym: Streptomyces avermectinius"  
complement(42. .2936)  
/gene="nrd"  
complement(42. .2936)  
/gene="nrd"  
/notes="SAV2461"  
/codon\_start=1  
/transl\_table=11  
/product="putative vitamin B12-dependent ribonucleotide reductase"

## gene

## CDS

/protein\_id="BAC70172.1"  
/db\_xref="GI:29606109"  
/translation="MTETASGPARSGRAKGTAKGLRIERIHTPTGVPVHVEVWAR  
VVTWNRDGSVNFQGVFPDFPWSVNAVIVTSKYFCGAVGTPOREVSRLQIDRI  
DYTKRAGEDIKYFASPADABIFEHAYALLAQIFSNFVFNVTGPQQVSACF  
ILAVDDSMESILDWYKEGMIFKGGAGLNLRSKLSKLLSGGNASGVPFMRGA  
DASAGTIKSGGATRAAKMVLIDVHPDIEDFTETKYKEEKIRALRDAGFMDLGG  
DITSVOYNNANSRVNDTFMAVEGGKFLSRMTGEVIEVDKSLFKMAEAAW  
ACADPGIOYDNTINHWHTCPESGRINGSNCPSEVHLDNTSCNLASLMLFKLDDGK  
GROSFEVERPAKVVELVITAMDISI CFADPTQKIGENTRAFGOLGIYANGLALMA  
TGHAYDSGGRALAGAITLMTGTSYKRSALAAVGVGYDGYARNAQPHQVQKQHS  
ANGAVRVDDLDTP IWAATAEWDVHLGKNGFRNAQSVIAPTGTIGLMSCDTT  
LPDLALVKFKLGVGSGI VNGTVFQALRLGYOEQIEAI VHAIAENGVIDAP  
GLKHEHYEVPDCAMGERSISAMGHVRMAAIQPIWISGALSKTLNLPETATVEDVEVY  
FEAWKGVKALAIYRNCVKQPILSAKKETAEVTAKEATIREAKVVEYRPPVR  
KRUPKRGPIGTTSTFGCAEGYNTFSPDDGGEVFLKMSKQSTLAGMWDASIAV  
SVGLQYQPLETYSKFTNNRFPAGMTDDPVRMAQSI VDIYIFRRALDLPETRS  
ALGHTSAEERQRLHETGSEYSDVDMDVQLAQSPRAQELKAVATPKAEVAAVPA  
PKQATSAELVEMQLQIADAPLCFCGCTKQKQAGSCYICGGCGSTGCS"  
complement(3086. .3649)  
/note="SAV2462"  
/codon\_start=1  
/transl\_table=11  
/product="putative regulatory protein"  
/protein\_id="BAC70173.1"  
/db\_xref="GI:29606110"  
/translation="MECPFCEHPDSRVVDSRTTDDGTSIRRRQPCDCSRRTTETC  
SLMVVKSGVTEPFRSTKTVNGVKACQRPVTEALAQICQVVEAVRATGSAELTT  
HDVGLALIFLRELDLVYLRFSVYFAFDSLEDEFAAIAELREQRPRVDDDEHE  
DAGAERQGTDRGSGGTVEVPATVAD"  
4168. .4947  
/gene="lexA"  
4168. .4947

## CDS

## gene

## CDS







Db	160952	CTGCTCAAGTCGACCGACCGGGCGCCACCTTGGGCGAGCGCGACCGCTTCCCG	160899	Qy	562	---PheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAla	580
Qy	217	GlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyVal	236	Db	159965	CCCTTCGAGGCCACCGCCGACATCGCCAGGAGCGACCGGACCGATCGCCACCGATGCC	159906
Db	160898	-----CGGAAGGCGAACTCTCTCGGCGAGGAGTC	160869	Qy	581	AspGlySerArgPheValTrpAla	596
Qy	237	ValTrpValAlaPheAspLysSerSerSerSerLeuGlyGlnAlaSerLysThrIlePhe	256	Db	159905	GACGCGCGCACACTGCTGTGTCCTTCTGTCATCGGACGCGC	159861
Db	160868	GTGTTTC-----CTCGTCGCGCGCGGCGCACCGTCTAC	160836	Qy	597	ValTyrAla	612
Qy	257	ValGlyValAlaAspProAsnAsnPro-----ValPheTrpSerArgAsp	271	Db	159860	ACGTACGCGCGCCACCGCTCGACGACAAACGCGCGAGCTGGTCCGAGGTCTCTCTCTTC	159801
Db	160835	GCCGCTGGGTGACGGCGACCGCACCTCGGGCACGGCGAACCTGTACCGCACGCGCGAC	160776	Qy	613	ProAlaAsnAlaGlnIleArgSerArgValAsnProLysThrPheTyrAla	630
Qy	272	GlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe-----IlePro	288	Db	159800	CCGAAGGGCGCACACCGGTGCGCGACCCGCGCATCCGACGGCTTCTACGCTACGAC	159741
Db	160775	--GGCAGCACTGGGGCGGTCCCGCGCGCGCTCCCGCACCTCCCGCAAGGTCCCG	160719	Qy	631	LeuSerAsnGlyThrPheTyrArgSerThrAspGlyValThrPheGlnProValAla	650
Qy	289	HisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGly	308	Db	159740	TTCGACAATGGAAACGCTATACGCCAGCACTGACAGTGGCGGTTCGTTTCACGCGCGGTGCG	159681
Db	160718	CTCCGCGCGCGTACGACACGACACCGCGAGCTGTACGTGACGTACGCGGACGACCC	160659	Qy	651	AlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysGlu	670
Qy	309	GlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrp	328	Db	159680	GGCGGACTGCCC---TCCGCGCACAGCCAGTTCAGCTGGTTCGCGCGCGCGGACGAAAGC	159624
Db	160658	GCCTCGGCGCGGCTCCGACGGCAGCGTGCACAGCTGCGTACCGCGCACGGGAGCTGG	160599	Qy	671	GlyAspLeuTrpLeuAlaAlaSer---SerGlyLeuTyrHisSerThrAsnGlyGlySer	689
Qy	329	ThrArgIleSerProVal-----ProSerThrAspThrAlaAsnAspTyrPhe	344	Db	159623	GGCGACCTGTGGCTCTCCGCCAAATGAAACGGGCTTACCGGTCCACCGACGGCGGAC	159564
Db	160598	ACCGAGGTACCCCGGTGAAGCGGGCGGACGACGAGCGAGCGGTCCGCGGACACCTTC	160539	Qy	690	SerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAla	709
Qy	345	GlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGln	364	Db	159563	ACCTTCGCCAGGATCGACAGCTGCTGGGCTCTGACACCTCGCTTCGCAAGCGGCC	159504
Db	160538	GCCTACGCGGGGTGCGCGTGCACCGCGCGCGCGCGCGCGCGTCTGCTCTCCACCAAC	160479	Qy	710	ProGlySerSerTyrProAlaValPheValValGlyThrIleGlyGlyValThrGlyAla	729
Qy	365	IleSerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThr	384	Db	159503	GACGCGCGCGATACCGCGCATCTACAGGTGGGTTCGACGGAGACCATCACCGCGCTC	159444
Db	160478	AACCGCTGGCGGACGCGGACACGCTCTTCGGGTCCAGGAGCGCGCGGTACCTGGAGC	160419	Qy	730	TyrArgSerAspAspCysGlyThrThrTrpValIleLeuAsnAspAspGlnHisGlnTyr	749
Qy	385	ArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSer	404	Db	159443	TACCGTCCGACGACCGCGCGGAGGATGGTCCGATCAACGACGACGCGCCACGATGG	159384
Db	160418	TCCTCAGGAC-----GCCGCGGTGTTCGACGTCTC	160386	Qy	750	GlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgValTyrIleGly	769
Qy	405	AlaGluProTrpLeuThrPheGlyValGlnProAsnProProValProSerProLysLeu	424	Db	159383	GGGTGATCGGAGCGCGCTCGTCTGATCCCGGATCCACGCGCGGTCTACTCTCGCC	159324
Db	160385	GAGACTCCCTCTCCAGCTGGGGCGACGAC-----AAGCCGAAGTTC	160344	Qy	770	ThrAsnGlyArgGlyIleValTyrGlyAspIle-----GlyGlyAlaPro	784
Qy	425	GlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGly	444	Db	159323	ACCAACGGCGCGGCATCCAGTACGGGAGCGGTCTGATGCCCGGCTCACGGATGCCA	159264
Db	160343	GGCTGGTGGATCCAGCGCTCGCGGTGACCCGTACGACTCCACGACGCTCGTGTACGGG	160284	Qy	785	SerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerProSerPro	804
Qy	445	ThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIle	464	Db	159263	CCCGGGCGCGCATCTCTGTTCGGTGGCG---ACTACAAACCCGAGACGTGGCCGAGGAGG	159207
Db	160283	ACCGGGCGACCTCTACGGCACCGCGACCTCAAGCGTGG-----	160242	Qy	805	Ser-----ProSerSerProSerPro-----	812
Qy	465	HisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAspLeuIleSerPro	484	Db	159206	TCTGGCAGGAGGACGTACGCTCATGAGGAGCGCGCGTCACTCGTCCGTCACCGTCGCGC	159147
Db	160241	-----GCACCGCGATCCGCGCTCGAGAGAGCGCGGTGCGCCAACTGATCTCGCCC	160188	Qy	813	SerProSerProSerSerPro-----	820
Qy	485	ProSerGly---AlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAla	503	Db	159146	TCTTCTCTGGGCGGAGCTCGAACCCACCCCGGGGCGAGGAGTTCGGTGGTTCGAC	159087
Db	160187	CCGTCGGGAGGCGCACCTGATACGCGACTCGGGGACATCGGTGTGATGTACACGAG	160128	Qy	821	-----SerSerSerProSerPro-----	828
Qy	504	AspValThrAlaValProSerThr---IlePheThrSerProValPheThrThrGlyThr	522	Db	159086	GGCTCATGGACCTGATGTCACGAGAACCGCATCGCGTCTGCTGGCCACCCCGCGCT	159027
Db	160127	CGGCTCACGGCTCTCTCGTCGGCGCATGGCGAAGAACCCCGGTTCGGGTGGCGGACG	160068	Qy	829	SerProSerPro-----	832
Qy	523	SerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspPro	542	Db	159026	CGCCCCCGCTCGATGGGCGGCTGTCACCCCGAGACCTTCCCGCGGACGAGGACGCGC	158967
Db	160067	GGACTCGCGCAGCGCGCGGCGGCGGTATGCTCGCACGGGCTGGGCGGACAC	160008	Qy	833	-----SerProSerArgSerProSerProSerAla	842
Qy	543	SerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrp---	561	Db	158966	GCATCGATGGTGGGGGCGCGGACGACCTTCGCCCACTCCAGTGCCTTACCGCGCT	158907
Db	160007	GGCAACGGC-----GGGTACTCCACGACGGCGCGGACCTGGCG	159966				







CDS	11570..13063			
	/gene="XAC1775"			
	/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"			
	/codon_start=1			
gene	/transl_table=11			
	/product="D-xylosidase"			
	/protein_id="AAM36639.1"			
	/db_xref="GI:21107973"			
CDS	/translation="MSLYVGLDVGTSVKLVAYDPQDRAVAVATIAAPMELISRDDGTR EQQAQWIDGIVHCFALQDGEORARVSGISVGGQHQGFVPAADGSVTPVKLWCDTST			
	Alignment Scores:			
	Pred. No.:	7,54e-27	Length:	14520
	Score:	1179.50	Matches:	312
Percent Similarity:	43.94%	Conservative:	134	
Best Local Similarity:	30.74%	Mismatches:	357	
Query Match:	22.97%	Indels:	213	
DB:	1	Gaps:	29	
CDS	US-09-917-376-1 (1-957) x AB011809 (1-14520)			
	Qy	27	AlaValAlaAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAlaHisAla	46
	Db	150	GCGTCTGCTGCGTGGTCTCTTTCATGACGTCCGAGGCATCCGCGGACGAGCCC	209
	Qy	47	AlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAsp	66
Db	210	GGTACGACAGACCGGTATCAGTGCACAGCGTGCATCGGGGGGGCGGCTTGTCA	269	
Qy	67	GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGly	86	
Db	270	GGTGTGCTGTTTCATCCCGCGAAGCGGATCTTGCTATATGACGACCGCATGTC	329	
Qy	87	MetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGly	106	
Db	330	GCTATCGCTGGGACGCGGACGAGCGTGGTGGCTGACCGATTTGTTGGGGGCC	389	
Qy	107	AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsn	126	
Db	390	GACGTTTGAATCTGATGGGAATCGACGCTTTTGGCGTTCGATCTCCGACCGA	449	
Qy	127	ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIle	146	
Db	450	TTGTACCTGGTGTGTCGACCTACATGCACGAACGCGCGGCAAT---GCCGCG	506	
Qy	147	ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGly	166	
Db	507	CGTTTCGTTCGATCCGGTAACAGATTCCGACCGCGGACCTGCCGTTCAAGTCG	566	
Qy	167	AsnMetProGlyArgGlyMetGlyValArgLeuAlaValAspProAsnAsnAsp	186	
Db	567	AATCAATTGGCGCGCCCAATGGCGAGCGCTGGCGTTGACCCGACGACGCGCGT	626	
Qy	187	LeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAla	206	
Db	627	CTGTTGCTTGCTCACGTGATGCG---GGCCTTGTGCGCAGCGACCATCGCGCG	683	
Qy	207	TrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAsp	226	
Db	684	TGGCGCGGGTGAAGGATTTCTCCCATGATGTTGGCTGGCGCCACCGCGCAACCA	743	
Qy	227	ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSer	246	
Db	744	GTCGGAACGGAACAGGCGGTG---GGCATCGCTTCTGTTGTTTCAGCGCGCGCG	800	
Qy	247	SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnPro	266	
Db	801	CATGCTGCGCTGCGACACCAACGATCATGTGTGGCGTTTCGACTCGCGCAAA	860	
Qy	267	PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGly	286	
Db	861	TATGTTCCGAAGATGCGGGGCGACCTGTTGCGGCGCTGCGCGGCAACCAAGG	920	











/gene="XCC1751"  
/note="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer/Genemark"  
/codon\_start=1  
/transl\_table=11  
/product="Tomb-dependent receptor"  
/protein\_id="AA041042.1"  
/db\_xref="GI:21112844"  
/translation="MMVGTLPYKKVEGFANAVNEVYDQVQWQITRPVNGDACKIRG  
AELGYTFDFLPWGLSGFLQNTYYVDSSEAPSTADTNGQLSLVPLEGLSKNSYN  
AILEYETARFGVAYNMRSDMLVTTAGNGTGNLPVYKNGFGQLDASLRFNINDVWSI  
SLDGNLLDTRRESYLGTSRYDRFVINDRYGLTLASL"  
7567..9822  
/gene="XCC1752"  
/db\_xref="XCC1752"  
7567..9822  
/gene="XCC1753"  
/note="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer/Genemark"  
/codon\_start=1  
/transl\_table=11  
/product="cellulase"  
/protein\_id="AA041043.1"  
/db\_xref="GI:21112845"  
/translation="WIMADIPPTKRRITTTGAWLAVLSLLLLLFTPNVRAAEPATSG  
PYORSAVAGGGFVGLPFAERGLAYARTDVGAVRMDAQAOQWALTDLGADG  
WNLGIDAFVADADAALYLAAGTYMHERAGNAVLRSFNRGTFERRADLPFLKGN  
QLRANGRLAVDPHGRVLLGSRDAGLWRRDGRHWKAVSFDDALAGATARNH  
VGRQAVFVFDVDSAGNTGTPTTPIYVGSSTEQSLYVSDAGRSWAPVAGPRG  
LRPMAGSGDGHVLYSGDQPDLMAGALKWFTPAQGRWREISPTPOPASGDGFG  
WGVAVDPOQVLLSTFRRTTPELRLYRSDGKHWAPLLADAVFDSAAPTAH  
TPHWGALIDPPGNHALFVTGYINASRLQDFAPQRPLOWFODRGLEETVPLD  
LLSPWAGHLLSALGDGDFRDELDRALQYAGPRLTNGESSIDAAGAPQWVVRSGT  
VDRNRNEILSDGDKQWTFASEPPAGGAGSIAIGADAAQVWAPERGNGWT  
SDFGQWRQDLPDPAVMDADRVDRRYAVDVASQQLYESTDAARSFRATGVQVGS  
PARBRTQPLRPDPFRAGVYVLLASPGKGVNRWDGTLQVLSQDEARSLGIGKALRA  
GAPALYLAGRVQGVGFRSDGQVQWRINDDAHRFRPVSVTGDPRIAGRVVFA  
GGRGIFYGDPB"  
9926..11839  
/gene="XCC1753"  
9926..11839  
/gene="XCC1753"  
/note="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer/Genemark"  
/codon\_start=1  
/transl\_table=11  
/product="sialic acid-specific 9-O-acetyl esterase"  
/protein\_id="AA041044.1"  
/db\_xref="GI:21112846"  
/translation="WTFSHISRACLLVLLSSGSAQAQPSLPLFADGAVLQREAMP  
VWGATPGASITVSDGQKATKDAQGWKVLIPAFHTAGGPPYVLSVHGDGGLQVRD  
VLVDGVLWLSQSNWEPQLAQSPPQAVANADQLRHFVKPSWSVOPETKLTGGD  
WKAAMPENAGAFSAVGFPAKELRSTGPIVDSITWGSAGIEMADMAISLGLKAEQ  
NQAIEATKORDAAQATQKRIGRWPKVDGTPQWRTPEFDRDWDISIPTYQWELS  
GYEGWDGIAWTFLLTAEEKAGVOLGVQIDSDITYVNGQVQVGTOKOMNLPV  
YAVPAALRAGVQIAVRVEDLSGCGMHGDEQRFVQIGSTQRLAGKFRPAQVRV  
SLTNKQPLTLLYNMHPQPPFVKGVIWYQGETNATDGTAKYRQDFAAMIQWR  
AERGAARLPLWQANFKAGDGLSPWALLRESQSKTLALPATGQAVIDIGNPT  
DIHPTNKGVGRHLAAHVAVGETLVYSAPFARASVADQASVQPDLOGSALQVR  
GGTTLQGFVAGADKRFHPAQAIRIAGDRVIVRSDAVAPVAVRYGWSNPADANLVR  
EQLVSPFRTDTW"

gene  
CDS

Alignment Scores:  
Pred. No.: 1.17e-26 Length: 11910  
Score: 1170.50 Matches: 299  
Percent Similarity: 48.61% Conservative: 138  
Best Local Similarity: 33.26% Mismatches: 350  
Query Match: 22.79% Indels: 113  
DB: 1 Gaps: 28

US-09-917-376-1 (1-957) x AE012276 (1-11910)

23 ThrAlaSerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSer 42

7606 ACCAACACAGAGCATGGCTGCTCTTGGAGCTGCTG---CTGTGCTCTTTTCGACG 7662  
43 Pro-----AlaHisAlaAlaThrGlnProTyrThrTyrSerAsnValAla 58  
7663 CCCAACGTGCTGGCGCGAGCCGCCACGCTCCGGGCGCTACACAGTGGCGAGTGTGCC 7722  
59 IleGlyGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeu 78  
7723 ATTGGCGGTGGCGGATTTGTACCGTGTGCTGTTTCATCCCGCGAACTGGTCTGGCC 7782  
79 TyrValArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIle 98  
7783 TATGCGGCACCGATGTGGTGGCGGTACCGCTGGATGCGAGCGCAGCAGTGGAC 7842  
99 ProLeuLeuAspTrpValGlyTyrAsnAsnTrpGlyTyrAsnGlyValValSerIleAla 118  
7843 GCGCTGACCGGACTGCTGGCGCGCCGACGACCTGGAACCTGATGGGATCGACGCAATTCGCC 7902  
119 AlaAspProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrp 138  
7903 GTCGACCCCGCGATGCGGATGCGCTGTATCTGGCGCGCGGACCTATATGACGACGACG 7962  
139 AspProAsnAspGlyValAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThr 158  
7963 GCCGCAAT---GCCGCACTGCTGCTGCTTCAACCGCGCGCGCAGCTTCGAGCGCGCC 8019  
159 ProLeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAla 178  
8020 GACCTGCGCTTCAAACTGGTGTAACTGAGTGGCGCGCAATGTCGCGAGCGCTGGCG 8079  
179 ValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrp 198  
8080 GTGGACCCACAGATGCGCGCGCTGCTGCTGGCGGCTCGCGCGATGCC---GGCCTGTGG 8136  
199 ArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThr 218  
8137 CGTAGCAGCATCGCGCGCGCTGGGCGAAAGTGGCTGCTTCCCGACGCGCGCTG 8196  
219 TyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrp 238  
8197 GCGGTGCGCACCGCACCGCAATCATGTTGGCGCGAGCAGCGGTG---GGGATCGCTTT 8253  
239 ValAlaPheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGly 258  
8254 GTCGTGTTTCGACCGCAGTGGCGCAACCGCGCACCGCAACCGCGCATCTACGTGGGC 8313  
259 ValAlaAspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAla 278  
8314 GTGTCCACCGAACACAGCAGCTGTATGTGTCGGAAGATGCGCGCGAGTGGGACCG 8373  
279 ValProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHis 298  
8374 GTGGCGGCGCAGCGCGCGCTGCGCGCGAGCCACATGCGCGCGCGCAGCATGGCAC 8433  
299 ValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspVal 318  
8434 ---TGTACTCTGAGCTACGCGCAGCAGCGCGCGCGGACCTGATGCGCGCGGAGCTTG 8490  
319 TrpLysPheSerValThrSerGlyThrTrpArgIleSerProValProSerThrAsp 338  
8491 TCAATAATTCACCGCGCAGCGCGCTGCGCGAGATAGCCCGGATTCGCGAG---CCA 8547  
339 ThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThr 358  
8548 GCCAGTGGCGGATTCGGCTGGGCGCGGTGGCGGTGATCCGCAACACCGCAGGTG 8607  
359 IleMetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAsp 378  
8608 CTGTTGGCGCAGCACCTTCCGCGCGCGCAGCGCGCGAGTGTATGCGAGCGGTGAT 8667  
379 GlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArg 398



Db 8668 GGTGGCAAGCACTGGCGCGCTG-----TTGGCCGAT 8700  
 Qy 399 TyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProPro 418  
 Db 8701 GCGGTGTTCGACCAAGCGCGCGCTGG---ACCGCACATGCCACCGCCAC 8751  
 Qy 419 ValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSer 438  
 Db 8752 -----TGGATGGG---GCGCTGGGATCGATCCGTTTCGACGCG 8787  
 Qy 439 AspArgMetLeuTrpGlyThrGlyAlaThrLeuTrpAlaThrAsnAspLeuThrLysTrp 458  
 Db 8788 AACACGCGCTGTCGTACCGCGCTACCGCATCTGGGCTCGCGCAATCTG 8838  
 Qy 459 AspSerGlyGlnIleHisIleAlaPro-----MetVal 470  
 Db 8839 -----CAGGATTCGCCGACCGCAGCGCGCGCTGCAGTGGTGTTCAGGAC 8886  
 Qy 471 LysGlyLeuGluGlyThrAlaValAsnAspIleSerProProSerGlyAlaProLeu 490  
 Db 8887 CGTGGCTTGAAGAAACCGTGCCTGACCTGCTCAGCCCGCATCGATCGCGCATCTG 8946  
 Qy 491 IleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSer 510  
 Db 8947 CTCAGCGCACTCGCGCATATCATGCTTCGGCATGACGATGGACCGCGCG---CAG 9003  
 Qy 511 ThrIlePheThrSerProValPheThrThrGlyThrSerValAspTrpAlaGluLeuAsn 530  
 Db 9004 TTGCAGTACCGCGCGCCCGCTGACCAATGCGAAGCATCGATGCGCGCGCGCGCGCG 9063  
 Qy 531 ProSerIleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHis 550  
 Db 9064 CGCAGTGGTGGTGGCGAGCGGTACCGTG---CGCGACCGCGCGCAACAAACGAAATCCGC 9120  
 Qy 551 ValAlaPheSerThrAspGlyGlyValAsnTrpPheGln---GlySerGluProGlyGly 569  
 Db 9121 GCGGTGTATTCGCTGATGGCGGCAAGCAATGACGGCGTTTGGCAGTGGAGCGCGCGCG 9180  
 Qy 570 ValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro 589  
 Db 9181 GGGCAGGCGCGCGCAGCATTCATGCTGGTGGCGATGCGCGCAGTGGTGGTGGCACCG 9240  
 Qy 590 GlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSer 609  
 Db 9241 GAGCGCGCGCGCAAT-----TGGCGCAGTCCGACTTCGCGCGCGCAGTGGCGAGCGTGTG 9294  
 Qy 610 GlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyr 629  
 Db 9295 GACGCGCTCGCGACACCGCGGTGGTGTGGCGATGCGGTGGACGACGCGCGCTGGTAT 9354  
 Qy 630 Ala-----LeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGln 647  
 Db 9355 GCGGTGGATGTCCGACGCGCGCAGCTGTACGAGAGCACCGATCCCGCGCGCATTTCCGT 9414  
 Qy 648 ProValAlaAlaGlyLeu----- 653  
 Db 9415 -----GCGACCGGTGTGAGTGGCGAGCGCGCGCGCGCGCGCGCGCGCGCA 9468  
 Qy 654 -----ProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysGluGly 671  
 Db 9469 CTGCGTCCGACCGCGCGCGCGCGCGGTGTATCTGGCCAGCGCGCGCGCGCGCGCG 9522  
 Qy 672 AspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerTrp 691  
 Db 9523 -----GGTGTGATCGCTGGCAGGACGCT-----ACGCTG 9552  
 Qy 692 SerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGly 711  
 Db 9553 CAGGTGCTCTCACAGCGCGCGAGCGCGCTGCTCGCATCGCGAAGCGCGCTGCGCGCG 9612  
 Qy 712 SerSerTyrProAlaValPheValGlyThrIleGlyGlyValThrGlyAlaTyrArg 731  
 Db 9613 GTGGCGCGCGCGCTGTATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9672

Qy 732 SerAspAspCysGlyThrThrTrpValLeuIleAsnAspAspGlnHisGlnTrpGlyAsn 751  
 Db 9673 TCCGACGATGGCGGTGTCAGTGGCAACCATCAACGATGACGCGCACCGCTTCGGCGCG 9732  
 Qy 752 TrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgValTrpIleGlyThrAsn 771  
 Db 9733 ---CGGTACAGCGGTACCGGTGATCCGCGCATTCGCGGCGGTGTGTACTTCGCCACCGCG 9789  
 Qy 772 GlyArgGlyIleValTrpGlyAspIleGlyAlaProSerGlySerProSerProSer 791  
 Db 9790 GCGCGCGCATTTCTACGCGCATCCGAGATGAGCAGTGCATGCCCTGCTCCGCA----- 9843  
 Qy 792 ValSerProSerAlaSerProSerLeuSerProSerProSerProSerProSerProSer 810  
 Db 9844 ---CGCCCTGCTGCTTACCGTGTACGGCGCGCGCGCGCGCTGGCGCGCGCGCGCGTGA 9900  
 Qy 811 -----SerProSerProSerProSerProSerProSerProSerProSerProSerProSer 827  
 Db 9901 CACCGCGCGCGCTGGAGATCACCGCATGACCCCGT---CCACATTTCCGCGCGCTGCTTGC 9959  
 Qy 828 -----ProSerProSerPro----- 832  
 Db 9960 TGGGT 10019  
 Qy 833 -----SerProSerArgSerProSerProSerProSerProSerProSerProSerProSer 846  
 Db 10020 CCGATGGCGCGGT 10079  
 Qy 847 SerSerSerProSerProSerProSerProSerProSerProSerProSerProSerProSer 865  
 Db 10080 GCGCTTCAA---TCAGGTGAGTTTCGATGGCAAGCAGGCGCACCGCGCACAGCCGACG 10133

Search completed: October 5, 2005, 16:22:25  
 Job time : 13428.3 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model  
Run on: October 5, 2005, 06:17:10 ; Search time 1472.27 Seconds  
(without alignments)  
3847.935 Million cell updates/sec

Title: US-09-917-376-1  
Perfect score: 5135  
Sequence: 1 MDRSENRLTRSRRLVSL.....RASFGSYNPATPTADTVLQX 957

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0  
Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cpn2.1/USPTO.spool/US09917376/runat 04102005.164327 26889/app query.fasta 1.2524  
-DB=N\_Geneseq\_16Dec04 -QWT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09917376 @CGN 1 1 1867 @runat 04102005.164327 26889 -NCPU=6 -ICPU=3  
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOF=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_16Dec04:.\*  
1: geneseqn1980s:.\*  
2: geneseqn1990s:.\*  
3: geneseqn2000s:.\*  
4: geneseqn2001as:.\*  
5: geneseqn2001bs:.\*  
6: geneseqn2002as:.\*  
7: geneseqn2002bs:.\*  
8: geneseqn2003as:.\*  
9: geneseqn2003bs:.\*  
10: geneseqn2003cs:.\*  
11: geneseqn2003ds:.\*  
12: geneseqn2004as:.\*  
13: geneseqn2004bs:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5134	100.0	2869	10 ADD22922	Add22922 Acidother
2	5134	100.0	2869	10 ABZ77632	Abz77632 Nucleotid
3	5130	99.9	2869	12 ADO52314	Ado52314 Acidother
4	2174	42.3	2823	8 ABV76941	Abv76941 Nucleotid
5	1631.5	31.8	2517	12 ADH51583	Adh51583 T reesei

6	1631.5	31.8	2710	12	ADH51580	Adh51580 Trichoder
7	1558.5	30.4	2849	10	ADD42060	Add42060 Trichoder
8	1167.5	22.7	3668	10	ADD24893	Add24893 DNA encod
9	1125.5	21.9	2646	12	ADH19117	Adh19117 Geotrichu
10	1124	21.9	2481	12	ADH19123	Adh19123 Geotrichu
11	1117	21.8	2367	12	ADH19119	Adh19119 Geotrichu
12	993.5	19.3	2755	13	ADR90295	Adr90295 Geotrichu
13	979	19.1	2268	13	ADR90297	Adr90297 Geotrichu
14	979	19.1	2274	13	ADR90301	Adr90301 Geotrichu
15	940	18.3	5698	10	ADD42054	Add42054 Trichoder
16	823	16.0	3687	8	ABZ77634	Abz77634 Nucleotid
17	823	16.0	3687	12	ADH36637	Adh36637 DNA encod
18	702	13.7	2289	8	ABZ77633	Abz77633 Nucleotid
19	702	13.7	2289	10	ADJ38292	Adj38292 A. cellul
20	429	8.4	3365	8	ABZ76162	Abz76162 A. cellul
21	409	8.0	11706	2	AAX55661	Aax55661 DNA seque
22	409	8.0	11707	6	AAD26525	Aad26525 Active ce
23	402	7.8	3364	10	ADF75896	Adf75896 Acidother
24	392.5	7.6	1103	3	AAF14988	Aaf14988 Trichoder
25	365.5	7.1	651	5	AAS84956	Aas84956 DNA encod
26	345	6.7	18596	4	AAF31109	Aaf31109 Thymidyla
27	345	6.7	18596	5	AAC91215	Aac91215 Human thy
28	345	6.7	18596	6	ABL67927	Ab167927 Ovary can
29	345	6.7	18596	6	ABL62854	Ab162854 Breast ca
30	345	6.7	18596	6	ABL63078	Ab163078 Breast ca
31	345	6.7	18596	6	ABK43334	Abk43334 Human Thy
32	345	6.7	18596	6	ABN95092	Abn95092 Gene #159
33	334.5	6.5	6415	2	AAX55662	Aax55662 DNA seque
34	334	6.5	6416	3	AAD26526	Aad26526 Active ce
35	330.5	6.4	7407	8	ACA42281	Acc42281 Prokaryot
36	330.5	6.4	7407	9	ACC59398	Acc59398 Microbial
37	323	6.3	2600	2	AAQ15178	Aaq15178 Portion o
38	319.5	6.2	8211	11	ABD15052	Abd15052 Pseudomon
39	318	6.2	1080	6	ABI99537	Abi99537 Mouse isc
40	302	5.9	1050	4	AAD21685	Aad21685 Mutaciona
41	302	5.9	2849	4	AAD21684	Aad21684 Human ret
42	299.5	5.8	210528	11	ACN44040	Acn44040 Mouse gen
43	296.5	5.8	4767	3	AAA14667	Aaa14667 Nucleotid
44	293	5.7	4818	3	AAA14669	Aaa14669 Nucleotid
45	290	5.6	1232	4	AAD11124	Adad11124 Human sma

ALIGNMENTS

RESULT 1  
ADD22922  
ID ADD22922 standard; DNA; 2869 BP.  
XX  
AC ADD22922;  
XX  
DT 15-JAN-2004 (first entry)  
XX  
DE Acidothermus cellulolyticus avicelase AvIII DNA.  
XX  
KW AvIII; cellulose reduction; agricultural biomass; municipal solid waste;  
KW Glycoside hydrolase; avicelase; ds; gene.  
XX  
OS Acidothermus cellulolyticus.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..2868  
FT /tag= a  
FT /product= "AvIII"  
XX  
PN US2003108988-A1.  
XX  
PD 12-JUN-2003.  
XX  
PF 18-OCT-2002; 2002US-00155400.  
PR 28-JUL-2001; 2001US-00917376.  
XX  
PA (DING/) DING S.



PA (ADNE/) ADNEY W S.  
PA (VINZ/) VINZANT T B.  
PA (HIMM/) HIMMEL M E.

PI Ding S, Adney WS, Vinzant TB, Himmel ME;

DR WPI; 2003-810853/76.  
DR P-PSDB: ADD22921.

AA  
PT New isolated thermal tolerant avicelase polynucleotide useful for  
PT detection of a polynucleotide encoding AvIII and for reducing cellulose  
PT in a starting material, e.g. municipal solid waste.

PS Claim 17; SEQ ID NO 2: 29pp; English.

The invention relates to an isolated polynucleotide molecule encoding a thermostable AviiiI polypeptide. The polynucleotide is useful for detection of a polynucleotide encoding AviiiI. The polynucleotide is useful for reducing cellulose in a starting material which involves administering to the starting material, e.g. agricultural biomass or municipal solid waste, a polypeptide molecule of the polynucleotide. The method further comprises administering a second polypeptide molecule chosen from the glycoside hydrolase family of proteins. The present sequence represents DNA encoding the *Acidothermus cellulolyticus* avicellase AviiiI.

Sequence 2869 BP: 546 A: 857 C: 900 G: 565 T: 0 U: 1 Other: XX

Alignment Scores:	1.46e-172	Length:	2869
Pred. No.:	Score:	Matches:	956
	5134.00	Conservative:	0
Percent Similarity:	100.00%	Mismatches:	0
Best Local Similarity:	100.00%	Indels:	0
Query Match:	99.98%	Gaps:	0
DP:	10		

US-09-917-376-1 (1-957) x ADD22922 (1-2869)

Qy	1	MetAspArgSerGluAsnIle	ArgLeuThrMetArgSerArgArgLeuValSerLeuLeu	20
Db	1	ATGGATCGTTCGGAGAAC	CATCCGTCGACTATGAGATCAGCAGATTTGGTATCACTGCTC	60
Qy	21	AlaAlaThrAlaSerPheAla	ValAlaAlaAlaLeuGlyValLeuProIleAlaIleThr	40
Db	61	GGCGCACATCGTTCGTTC	CGCGCCGCTCTGGGAGTTCTGCCCATCGCGATAACG	120
Qy	41	AlaSerProAlaHisAla	AlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGly	60
Db	121	GCTTTCCTGCGCAGCGCG	GACGACTCAGCCGTACCTGGAGCAACATGGCGATCGGG	180
Qy	61	GlyGlyGlyPheValAspGly	IleValPheAsnGluGlyAlaProGlyIleLeuTyrVal	80
Db	181	GGCGCGGCTTTGTGCGA	CGGGATCGTCTTCAATGAAGTGCACCGGAATCTGTACGTG	240
Qy	81	ArgThrAspIleGlyGlyMet	TyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeu	100
Db	241	CGGACGCAATCGGCGGGAT	GTATCGATGGGATGCGGCCAACCGGCGGTGGATCCCTCTT	300
Qy	101	LeuAspTrpValGlyTrpAsn	AsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAsp	120
Db	301	CTGGATTTGGGTGGGATGG	AACAAATTTGGGGGTACAAACGGCGTCGCAGCATTTGGCGCAGAC	360
Qy	121	ProIleAsnThrAsnLysVal	TrpAlaAlaValGlyMetTyrThrAsnSerTrpAspPro	140
Db	361	CCGATCMAATACTAACAG	GTATGGGCCCGCGTCGGATGTACACCAACAGCTGGGACCCA	420
Qy	141	AsnAspGlyAlaIleLeuArg	SerSerAspGlnGlyAlaThrTrpGlnIleThrProLeu	160
Db	421	AACGACGGACGATTTCTC	CGCTCGTCTGATCAGGGCGCAACGTGCGCAAAATAACGCCCTG	480
Qy	161	ProPheLysLeuGlyLysAsn	MetProGlyArgGlyMetGlyGluArgLeuAlaValAsp	180
Db	481	CCGTTCAACGCTTGGCGGCA	ATGCGCGCGCTGGAAATGGCGAGCGGCTTGGCGTGGAT	540



QY 541 AspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLeuAsn 560  
DB 1621 GATCATCGAGCCCAACCGAAGCAGGACGTCGCGTCTCGACAGACGGCGGCAAGAAC 1680  
QY 561 TrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAla 580  
DB 1681 TGGTTCNAGCAGCAGCACTGGCGGGGTGACGACGGCGGCACCGTCGCGCATCGGCC 1740  
QY 581 AspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValThrAlaVal 600  
DB 1741 GACGGCTCTCGTTCGTCGGCTCCCGCGCATCCCGGTGAGCCCTGTGTGTAGCAGTC 1800  
QY 601 GlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAlaGlnIleArgSer 620  
DB 1801 GGATTGGCAACTCTCGGGCTGCTTCGCAAGGTGTTCGCCCAATGCCAGATCCGCTCA 1860  
QY 621 AspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThr 640  
DB 1861 GACCGGGTGAATCCAAAGACTTCTATGCCCTTCCATGGAACTTCTATCGAAGCAG 1920  
QY 641 AspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGly 660  
DB 1921 GACGCGCGGTGACATTCACCGCGTCCGCGCGGTCTTCCGAGCAGCGGTGCGCTCGGT 1980  
QY 661 ValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGly 680  
DB 1981 GTCATGTTCCACGCGGTGCTGGAAAGAGGCGATCTGTGCTCGCTGCATCGAGCGGG 2040  
QY 681 LeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAla 700  
DB 2041 CTTTACCACCTCAACCAATGGCGGAGCAGTGGTCTGCATCACCAGCGGTCTTTGTGCTC 2100  
QY 701 ValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValVal 720  
DB 2101 GTGAACGTGGGATTTGGTAAGTCTGCGCGCGGTGCTCATACCCAGCGCTTTGTGCTC 2160  
QY 721 GlyThrIleGlyGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpVal 740  
DB 2161 GGCACGATCGGAGCGGTTACGGGGCGGTACCGCTCCGACGACTGTGGACGACCTGGGTA 2220  
QY 741 LeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHis 760  
DB 2221 CTGATCAATGATGACGAGCACCACCAATACGGAATTTGGGACAAAGCAATACCGGTGACAC 2280  
QY 761 AlaAsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle 780  
DB 2281 GCGAATTTACGGGGGTGTACATAGGCACGACGCGCGTGGAAATGTATACGGGACATT 2340  
QY 781 GlyGlyAlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeu 800  
DB 2341 GGTGTGCGCGGTTCGGATCGCGCTCTCGGTGAGTCCGTTCGCTTCGCGGACGCTG 2400  
QY 801 SerProSerProSerProSerProSerProSerProSerProSerProSerProSerPro 820  
DB 2401 AGCCCGAGCCGAGCCGAGCTCGCCATCGCGTCCGCGTCCGCGGAGCTCGAGTCCA 2460  
QY 821 SerSerSerProSerProSerProSerProSerProSerProSerProSerProSerPro 840  
DB 2461 TCTCTGTCGCGGTCTCCGTCGCGGTCCATCCATCCGAGTCCGCTTCGCTTCGCTCACCA 2520  
QY 841 SerAlaSerProSerProSerProSerProSerProSerProSerProSerProSerPro 860  
DB 2521 TCGCGGTCCGCGGCGCGCTTCGTCACCGAGCCGCTCTCGTACCGCTTCGTCGCGCG 2580  
QY 861 SerProThrProSerSerSerProValSerGlyValIleValGlnTyrIleAsnAsn 880  
DB 2581 AGCCCAACGCGCTCGTTCGCGCGGTGTCGGGTGAGGTGAGGTATAGAAATAAT 2640  
QY 881 AspSerAlaProGlyAspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySer 900  
DB 2641 GATTCGCGCGCGGTGATTAATCAAGATCAAGCCGGGTTTCAGGTGTGTGAATACCGGGTGC 2700  
QY 901 SerSerValAspLeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySer 920

DB 2701 TCGTCGGTGGATTTGTCGACCGGTGACCGGTGCGGTACTGTTTCCCGGATGCTGCTCG 2760  
QY 921 SerThrLeuValTyrAsnCysAspTrpAlaIleGlyCysGlyAsnIleArgAlaSer 940  
DB 2761 TCGACACTGGGTGTACACTGTGACTGGCGGGGATCGGGTGTGGAAATATCCGCGCTCG 2820  
QY 941 PheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 956  
DB 2821 TTGCGCTCGTGTGAACCGCGGACCGGCGGACACCTACTCTGCAG 2868  
RESULT 2  
ID ABZ77632 standard; DNA; 2869 BP.  
XX AC ABZ77632;  
XX 03-JUN-2003 (first entry)  
XX Nucleotide sequence of the avicelase AviIII.  
XX Avicelase; AviIII; glycoside hydrolase; enzyme; cellulase; biofuel;  
KW detergent; pulp processing; paper processing; feed processing; textile;  
KW cellulose; gene; ss.  
XX Acidothermus cellulolyticus.  
XX FH Location/Qualifiers  
FT CDS 1..2869  
FT /\*tag= a  
FT /partial  
FT /product= "AviIII"  
FT /transl\_except= (pos:2869,aa:Xaa)  
FT /note= "Xaa is an unspecified residue"  
XX PN WO2003012090-A2.  
XX PD 13-FEB-2003.  
XX PF 28-JUL-2001; 2001WO-US023818.  
XX PR 28-JUL-2001; 2001WO-US023818.  
XX PA (MIDE ) MIDWEST RES INST.  
XX PI Ding S, Adney WS, Vinzant TB, Himmel ME;  
XX WPI: 2003-248177/24.  
XX P-PSDB; ABP73015.  
XX New thermostable AviIII peptide from Acidothermus cellulolyticus, useful  
PT for degradation of cellulose or in generating anti-AviIII antibodies for  
PT purifying recombinant AviIII polypeptides from genetically engineered  
PT host cells.  
XX Claim 3; Page 24; 44pp; English.  
XX The present sequence encodes a thermostable avicelase polypeptide,  
CC designated AviIII. AviIII is a member of the glycoside hydrolase family  
CC of enzymes, and is a cellulase. AviIII is useful in the conversion of  
CC biomass to biofuels and biofuel additives. It may be useful in the  
CC production of detergents, pulp and paper processing, food and feed  
CC processing and in textile processes. The thermostable AviIII peptide is  
CC useful in the degradation of cellulose, and in generating specific anti-  
CC AviIII antibodies that are useful in purifying recombinant AviIII  
CC polypeptides from genetically engineered host cells, in detecting AviIII  
CC polypeptide expression, as well as a reagent tool for characterizing the  
CC molecular actions of the polypeptide. The AviIII polynucleotide is useful  
XX as a source of probes or primers in various diagnostic assays  
XX Sequence 2869 BP; 546 A; 857 C; 900 G; 565 T; 0 U; 1 Other;  
Alignment Scores:



Pred. No.:	1.46e-172	Length:	2869
Score:	5134.00	Matches:	956
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	99.98%	Indels:	0
DB:	10	Gaps:	0
US-09-917-376-1 (1-957) x ABZ77632 (1-2869)			
QY	1	MetAspArgSerGluAsnIleArgLeuThrMetArgSerArgArgLeuValSerLeuLeu	20
DB	1	ATGGATCGTTCGGAGAAATCCGTCGTGACTATGAGATCAGCAGGATGGTATCACTGCTC	60
QY	21	AlaAlaThrAlaSerPheAlaValAlaAlaAlaLeuGlyValLeuProIleAlaIleThr	40
DB	61	GCCGCCACTGGTCGTTCCCGTGCCGCCCGCTCTGGAGATTCTGCCCATCGCGATAAG	120
QY	41	AlaSerProAlaHisAlaAlaThrGlnProIleProIleThrTrpSerAsnValAlaIleGly	60
DB	121	GCTTCTCTCGCGACGCGCGAGACTCAGCGGTACACTGGAGCAACGTTGGCGATCGG	180
QY	61	GlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuThrVal	80
DB	181	GCGCGCGCTTGTGCACGGGATCGTCTCAATGAAGGTGCACCGGGAATTCGTACGTG	240
QY	81	ArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyValGlyTrpIleProLeu	100
DB	241	CGACGGACATCGGGGGATGATCGATGGGATGCCCAACGGCGGTGGATCCCTCTT	300
QY	101	LeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAsp	120
DB	301	CTGGATTGGTGGATGGACAAATTTGGGGGTACACGGCGTGTGATTCGCGCAGAC	360
QY	121	ProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspPro	140
DB	361	CCGATCAATACTAAACAAGGTATGGCGCCCGCTCGGAATGTACACCAACAGCTGGGACCCA	420
QY	141	AsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeu	160
DB	421	AACGACGGAGGAGATTCCTCGCTCGTCTGATCAGGGCGCAACGTGGCAATAACGCCCTG	480
QY	161	ProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAsp	180
DB	481	CGTTCAAGCTTGGCGGCAACATGCCCGGGGTGAATGGGCGAGCGGCTTCGGTGGAT	540
QY	181	ProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSer	200
DB	541	CCAAACAATGACAACTTCTGATTTTCGGCGCCCGCGAGCGGCAAGGGGCTCTGGAGAAGC	600
QY	201	ThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIle	220
DB	601	ACAGATTCGGCGGACCTGGTCCAGATGACGAATCTTTCGGGACGTAGGCACGTACATT	660
QY	221	AlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAla	240
DB	661	GCAATCCCACTGACACGCCGCTATCAGACGATATTCAGGCGTGTCTGGGTGCGT	720
QY	241	PheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAla	260
DB	721	TTCCGACAAGTCTTCGTATCGCTCGGCAAGCGAGTAAGACCAATTTTGTGGCGTGGCG	780
QY	261	AspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValPro	280
DB	781	GATCCCAATAATCCGGTCTCTTCGGAGCAGACGCGCGCGACGTGGCAGCGGTGGCG	840
QY	281	GlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeu	300
DB	841	GGTGGCGGACCGGCTTCAATCCCGCACAGGGCGCTTTTGACCCGCTCAACCACGTGCTC	900
QY	301	TyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLys	320
DB	901	TATATTGCCACCAGCAATACGGTGGTCCGTATGACGGGAGTCCCGCGACGCTCCGGCGACGCTCTGGAAA	960
QY	321	PheSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAspThrAla	340
DB	961	TTTCGGTGACCTCCGGGACATGGACGGAAATCAGCCCGGTACCTTCGACGGACACGGCC	1020
QY	341	AsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMet	360
DB	1021	AACGACTACTTTGGTTACAGCGGCTCTACTATCGACCGCGCAGCACCAGCAATG	1080
QY	361	ValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGly	380
DB	1081	GTGGCAACCCAGATATCTGGTGGCGGACACATAATCTTTCGGAGACCGGACGGCGT	1140
QY	381	AlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrVal	400
DB	1141	GGCAGTGGACGGGATCTGGGATTTGGACAGTATCCCAATCGAAGCTTCGCATATGTG	1200
QY	401	LeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProProValPro	420
DB	1201	CTTGACATTTCCGGCGAGCTTTGGCTGACCTTCGGCGTACAGCCGAATCTCCCGTACCC	1260
QY	421	SerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg	440
DB	1261	AGTCGAAGCTCGCTGGATGGATGAAGCGATGGCAATCGATCCGTTCACTCTGATCGG	1320
QY	441	MetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSer	460
DB	1321	ATGCTCTACGGAACAGGCGCGAGCTTGTACGCAACAAATGATCTCACGAAGTGGGACTCC	1380
QY	461	GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAsp	480
DB	1381	GGCGGCGAGATTCATATCGCGCGGATGGTCAAAAGGATTTGGAGAGACGGCGTAAACGAT	1440
QY	481	LeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPhe	500
DB	1441	CTCATCAGCCCGCGCTCTGGCGCCCGCTCATCAGCGCTCTCGAGACCTTCGGCGGCTTC	1500
QY	501	ThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThr	520
DB	1501	ACCCAGCGGAGCTTACTGCGGTGCCATCGACGATCTTCACGTCACCGGTGTTCACGACC	1560
QY	521	GlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPhe	540
DB	1561	GGCACCGCGTCACTATCGGAAATTTGAATCGCTGCATCATCTGTCGCGTGGAAAGTTTC	1620
QY	541	AspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsn	560
DB	1621	GATCATTCAGGCAACCCGAAACACAGGACGTCGCGTCTTCGACACAGCGCGGCAAGAC	1680
QY	561	TrpPheGlnGlySerGluProGlyValThrThrGlyGlyThrValAlaAlaSerAla	580
DB	1681	TGGTTCCAAGGACGACGACCTGGCGGGGTGACGACGGGCGGACCGTCGCGCATCGGCC	1740
QY	581	AspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaVal	600
DB	1741	GACGGCTCTCGTTTCTGTCGGCTCCCGCGCATCCCGGTGAGCTGTGTGTACGCAATC	1800
QY	601	GlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSer	620
DB	1801	GGATTTGGCAACTCTCTGGGCTGCTTCGCAAGGTGTTCCCGCAATCCCAAGATCCCGTCA	1860
QY	621	AspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThr	640
DB	1861	GACCGGGTGAATCCAAAGACTTTCTATGCTTATCCATATGGAACCTTCTATCGAAGCAG	1920
QY	641	AspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGly	660
DB	1921	GACGGCGGCTGACATTCACCGGTCCGCGCGGTCTTCCGAGCAGCGGTGCGTCCGT	1980
QY	661	ValMetPheHisAlaValProGlyLysGlyGlyAspLeuTrpLeuAlaSerSerGly	680
DB	1981	GTCAATTTCCACGCGGTGCTGGAAAAGAGGATCTGTGGTCTGCTGATCGAGCGGG	2040
QY	681	LeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAla	700



Db 2041 CTTTACCACTCAACCAATGGCGGACGAGTGGTCTGCAATCACCGGATCTCTCCCG 2100  
 Qy 701 ValAenValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValVal 720  
 Db 2101 GTGAAGCTGGGATTTGGTAAAGTCTGCGCCCGGCTCGTATACCCAGCGCTCTTGTCGTC 2160  
 Qy 721 GlyThrIleGlyGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpVal 740  
 Db 2161 GGCACGATCGGAGCGGTTACGGGGCGGTACCGCTCCGACGACTGTGGGACGACCTCGGTA 2220  
 Qy 741 LeuIleAenAspAspGlnHisGlnTyrGlyAenTyrGlyGlnAlaIleThrGlyAspHis 760  
 Db 2221 CTGATCAATGATGACACGACCAATACGGAATTTGGGACAGCAATACCGGTGACCCAC 2280  
 Qy 761 AlaAenLeuArgArgValTyrIleGlyThrAenGlyArgGlyIleValTyrGlyAspIle 780  
 Db 2281 GCGAATTTACGGCGGTTGTACATAGCAGCAAGCGCGTGGAAATTTGTATACGGGACATT 2340  
 Qy 781 GlyGlyAlaProSerGlySerProSerProSerProSerProSerProSerProSerPro 800  
 Db 2341 GGTGTGCGCGCTCGCGATCGCGTCTCGTCTCGTGTAGTCTCGTCTCGCGACCTG 2400  
 Qy 801 SerProSerProSerProSerProSerProSerProSerProSerProSerProSerPro 820  
 Db 2401 AGCCCCGAGCCGAGCCGAGCGAGCTCGCCATCGCCGTGCGCTCGCGAGCTCGAGTCA 2460  
 Qy 821 SerSerProSerProSerProSerProSerProSerProSerProSerProSerProSer 840  
 Db 2461 TCTGTCTGCGCTCTCGCTCGCGCTCACCATCGCCGAGTCTCGTCTCGCTCACCA 2520  
 Qy 841 SerAlaSerProSerProSerProSerProSerProSerProSerProSerProSerPro 860  
 Db 2521 TCGCGCTCGCGAGCCGCTCTCGTCAAGCGCGCTCTCGTCAAGCGCTCTCGTCTCG 2580  
 Qy 861 SerProThrProSerProSerProSerProSerProSerProSerProSerProSerPro 880  
 Db 2581 AGCCCAACCGCTCGTCTCGCGCTGCGGTGGGTGAGTGTGAGTGTGAGTGTGAGT 2640  
 Qy 881 AspSerAlaProGlyAspAenGlnIleLysProGlyLeuGlnValValAenThrGlySer 900  
 Db 2641 GATTTCGGCGCGGGTGATTAATCAGATCAAGCCGGTTTCAGGTGGTGAATACCGGTG 2700  
 Qy 901 SerSerValAspLeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySer 920  
 Db 2701 TCGTGTGTGATTTGTCGAGCGGTGACGCTGCGTACTGTTTCAAGCGGATGCTGCTCG 2760  
 Qy 921 SerThrLeuValTyrAenCysAspTrpAlaAlaIleGlyCysGlyAenIleArgAlaSer 940  
 Db 2761 TCGACACTGTGTACAACTGTGACTGGCGCGGATCGGGTGGGAATATTCGCGGCTCG 2820  
 Qy 941 PheGlySerValAenProAlaThrProThrAlaAspThrTyrLeuGln 956  
 Db 2821 TTGCGTCTGCTGAACCCGCGAGCCGCGACGCGGACACCTACTCTGCGAG 2868

RESULT 3

ID ADO52314 standard; DNA; 2869 BP.

XX AC ADO52314;

XX DT 15-JUL-2004 (first entry)

XX AC Acidothermus cellulolyticus avicelase III (AviIII) DNA.

XX DE Thermostable cellulase; avicelase III; AviIII; cellulose reduction;  
 KW biomass degradation; ethanol formation; industrial chemical;  
 KW fabric treatment; gene; ds.  
 XX OS Acidothermus cellulolyticus.

XX FH Key Location/Qualifiers

FT CDS 1..2868

FT FT /\*tag= b  
 FT /product= "Acidothermus cellulolyticus avicelase III"  
 FT /note= "CDS contains translational exceptions"  
 FT /partial  
 FT /note= "No stop codon"  
 FT 1..306  
 FT sig\_peptide /\*tag= a  
 FT 307..2865  
 FT mat\_peptide /\*tag= c  
 FT /product= "Acidothermus cellulolyticus mature avicelase  
 FT III"  
 FT 2869  
 FT misc\_feature /\*tag= d  
 FT /note= "A string of unknown nucleic acid units"  
 FT XX  
 PN US2004038334-A1.  
 XX 26-FEB-2004.  
 XX 28-JUL-2001; 2001US-00917376.  
 XX 28-JUL-2001; 2001US-00917376.  
 XX (DING/) DING S.  
 PA (ADNE/) ADNEY W S.  
 PA (VINZ/) VINZANT T B.  
 PA (HIMM/) HIMMEL M E.  
 XX Ding S, Adney WS, Vinzant TB, Himmel ME;  
 PI WPI: 2004-203224/19.  
 DR P-PSDB; ADO52313.  
 DR Novel thermostable AviIII polypeptide of glycoside hydrolase family and  
 PT isolated from Acidothermus cellulolyticus, useful for degrading treated  
 PT biomass into simpler forms of carbohydrate.  
 XX Claim 17; SEQ ID NO 2; 19pp; English.  
 XX The invention relates to a thermostable cellulase enzyme, avicelase III  
 CC (AviIII) and its nucleic acid sequence. AviIII is useful for reducing  
 CC cellulose in a starting material. A thermostable AviIII peptide is useful  
 CC for degrading treated biomass into simpler forms of carbohydrate, which  
 CC is used in the formation of ethanol or other industrial chemicals. It is  
 CC also useful for treating fabrics to remove cellulose-containing stains.  
 CC The present sequence is Acidothermus cellulolyticus AviIII DNA.  
 XX SQ Sequence 2869 BP; 546 A; 857 C; 899 G; 565 T; 0 U; 2 Other;  
 Alignment Scores:  
 Pred. No.: 2,02e-172 Length: 2869  
 Score: 5130.00 Matches: 955  
 Percent Similarity: 99.90% Conservative: 0  
 Best Local Similarity: 99.90% Mismatches: 1  
 Query Match: 99.90% Indels: 0  
 DB: 12 Gaps: 0  
 US-09-917-376-1 (1-957) x ADO52314 (1-2869)  
 Qy 1 MetAspArgSerGluAenIleArgLeuThrMetArgSerArgArgLeuValSerLeuLeu 20  
 Db 1 ATGGATCGTTCCGGAGAACATCCGCTGACTATGAGTACACGACGATGTTGATCACTGTC 60  
 Qy 21 AlaAlaThrAlaSerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThr 40  
 Db 61 GCGCCCACTCGGTCTGCGCGTGGCGCGCTCTGGGAGTCTTCCCATCGGATAACG 120  
 Qy 41 AlaSerProAlaHisAlaAlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGly 60  
 Db 121 GCTTCTCTCGCGACGCGCGGACGACTCAGCCGTACACCTCGGAGCAACGTCGCGATCGGG 180  
 Qy 61 GlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrVal 80



Db 181 GCGCGCGCTTTGTTCGACGGGATCGTCTTCAATGAAGGTGCACCGGGAAATTTCTGTACCTG 240  
Qy ArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaAenGlyValTrpIleProLeu 100  
Db 241 CGGACGGACATCGGGGGGATGATTCGATGGGATGCGCCCAACGGGCGGTGGATCCCTCTT 300  
Qy 101 LeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAenGlyValValSerIleAlaAlaAsp 120  
Db 301 CTGGATTGGGTGGATGGAAACAATGGGGGTACAAACGGCGTCGTTCAGCATTTGGCGCAGAC 360  
Qy 121 ProfileAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspPro 140  
Db 361 CCGATCAATACTAAACAGGTATGGCGCGCTCGGAATGTACACCAACAGCTGGGACCCA 420  
Qy 141 AsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeu 160  
Db 421 AACGACGGAGGATTTCTCGCTCGTTCATCAGGGCGCAAGGTGGCAATTAACGGCCCTG 480  
Qy 161 ProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAsp 180  
Db 481 CCGTTCAAGCTTGGCGGCAACATGCCCGGGGTGGATGGGCGAGCGCTTGGCGTGGAT 540  
Qy 181 ProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSer 200  
Db 541 CCAACAATGACAAATTTCTGATTTTCGGCGCCCGAGCGGCAAAAGGCTCTGGAGAAAGC 600  
Qy 201 ThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIle 220  
Db 601 ACAGATTCGGCGCGACCTTGGTCCAGATGACGAATTTTCGGGACGTAGGCACGTACATT 660  
Qy 221 AlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAla 240  
Db 661 GCAATCCCACTGACACACCGCGCTATCAGAGCGATATTCAGGCGTCTGTGGGTGCGT 720  
Qy 241 PheAspLysSerSerSerLeuGlnAlaSerLysThrIlePheValGlyValAla 260  
Db 721 TTTCGACAAAGTCTTCGTATCGCTCGGCAAGCGAGTAAGACCATTTTGTGGCGTGGCG 780  
Qy 261 AspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValPro 280  
Db 781 GATCCCAATAATTCGGGTCTTCGAGACAGACGCGCGCGACGTGGCGCGGTGGCG 840  
Qy 281 GlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeu 300  
Db 841 GGTGGCGCGACCGGCTTATCCCGCACAGGCGCTTTTGACCCCGGTCAACCACTGTCTC 900  
Qy 301 TyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLys 320  
Db 901 TATATTGCCACCAGCAATACGGGTGTCGTATGACGGGAGCTCCGGCGACGTCTTGAAA 960  
Qy 321 PheSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAspThrAla 340  
Db 961 TTTCGCGTGACCTCCGGGACATGAGCGGAATTCAGCCCGGTACCTTCGACGGACACGGCC 1020  
Qy 341 AsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMet 360  
Db 1021 AACGACTACTTTGGTTACAGCGGCTCACTATCGACCGCGCACACCCGAAACAGATAATG 1080  
Qy 361 ValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGly 380  
Db 1081 GTGGCAACCCAGATATCGTGGTGGCGGACCACTAATCTTTCGGAGCACCCACGGCGGT 1140  
Qy 381 AlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrVal 400  
Db 1141 GCGACGTGGACCGCGATCTGGGATTTGGACGAGTTATCCCAATCGAAGCTTGGGATATGTG 1200  
Qy 401 LeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProProValPro 420  
Db 1201 CTTGACATTTTCGGCGAGCCTTGGGTGACCTTCGGCGGTACACCGGAATCCCTCCCGTACCC 1260  
Qy 421 SerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg 440  
Db 1261 AGTCCGAAGCTCGGTGGATGGATGAGCGGATGGCAATCGATCCGTTCACTCTGATCGG 1320

Qy 441 MetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSer 460  
Db 1321 ATGCTCTACGGACACAGCGCGAGCTGTGTACGCAACAATGATCTCACGAAGTGGACTCC 1380  
Qy 461 GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAsp 480  
Db 1381 GCGCGCCAGATTCATATCGCGCGATGGTCAAAAGGATTCGAGGAGACGCGGTAAACGAT 1440  
Qy 481 LeuIleSerProProSerGlyAlaProLeuIleSerAlaIleGlyAspLeuGlyGlyPhe 500  
Db 1441 CTCATCAGCCCGCTCTGGCGCCCGCTCATCAGCGCTCTTCGAGACCTTCGGCGGCTTC 1500  
Qy 501 ThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThr 520  
Db 1501 ACCACGCGAGCTTACTTGGCGGTGCATCGACGATCTTCACGTACCGGTGTTCACGACC 1560  
Qy 521 GlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPhe 540  
Db 1561 GGCACCAGCGTCGACTATCGGGAATTTGAATCCGTGCATCATCGTTCGCGCTGGAAGTTTC 1620  
Qy 541 AspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsn 560  
Db 1621 GATCCATCGAGCAACCCGAACGACAGCGACGTGCGCTTCTCGACACGCGCGCAAGAAC 1680  
Qy 561 TrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAla 580  
Db 1681 TGGTTCGAAGGACGACCACTGGCGGGGTGACACGCGGGGACCGTTCGCGCATCGGCC 1740  
Qy 581 AspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaVal 600  
Db 1741 GACGCTCTCTGTTTGTCTGGGCTCCCGCGCATCCCGGTTCAGCTGTGTGTACGCAATC 1800  
Qy 601 GlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSer 620  
Db 1801 GGATTTGGCAACTCTCTGGGCTGCTTCGCAAGGFGTTCCCGCCAATCCCAAGATCCCGTCA 1860  
Qy 621 AspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThr 640  
Db 1861 GACCGGGTGAATCCAAAGACTTTTCTATGCCCTATCCATGGAACCTTCTATCCAGACAG 1920  
Qy 641 AspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGly 660  
Db 1921 GACGGCGCGTGCATTCACACCGGTTCGCGCGCGCTTCCTCGAGACGCGGTGCGTGGT 1980  
Qy 661 ValMetPheHisAlaValProGlyLysGlyGlyAspLeuTrpLeuAlaAlaSerSerGly 680  
Db 1981 GTCATGTTCCACGCGTGCCTGGAAAAGAAAGGCGATCTGTGGCTCGTGCATCGAGCGGG 2040  
Qy 681 LeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAla 700  
Db 2041 CTTTACCACCTCAACCAANTGGCGCAGAGTTGGTCTGCAATACCCGCGGTATCTCTCCGG 2100  
Qy 701 ValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValVal 720  
Db 2101 GTGACGTGGGATTTGGTAAGTCTGCGCCCGGCTCGTCATACCACGCGCTTTTGTGCTC 2160  
Qy 721 GlyThrIleGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpVal 740  
Db 2161 GGCACGATCGGAGCGCTTACGGGGCGGTACCGCTCCGACGACTGTGTGGACGACCTGGGTA 2220  
Qy 741 LeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHis 760  
Db 2221 CTGATCAATGATGACAGCACCAATACGGAATTTGGGGAACAACAATACACCGGTGACCA 2280  
Qy 761 AlaAsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle 780  
Db 2281 GCGAATTTACGGCGGGGTGTACATAGGACGAAACGGCGGTGGAAATTTGTATACGGGGACAT 2340  
Qy 781 GlyGlyAlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeu 800  
Db 2341 GGTGGTGGCGCTCCGGATCGCGCTCTCCGTCCGTGAGTCCGTCCGCTTCCCGAGCGCTG 2400







QY 249 GlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrp 268  
DB 715 GGTACAGACAAACCAACCATCTACGTAGGTGTAGCAGACAGCAAGCAACAGTTTATCGA 774  
QY 269 SerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIlePro 288  
DB 775 TCTACCGATGGGGGCAACATGCGCAACGCGTCCCTGGACACACCCACCGGATTCCTGGCG 834  
QY 289 HisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGly 308  
DB 835 CAAAGAGGAGTCTTGACCATAAAGGTGAGCAACTCTACATCGCTACCTCAGATACGTGT 894  
QY 309 GlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrp 328  
DB 895 GCCCGTATGACGGGTGAGGGGATGTGTGGCCCTAGATATTTCCAGTGGCCAGTGG 954  
QY 329 ThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGly 348  
DB 955 ACTCGAATCAGCCCAATTCCTCTACC--TCATCGAATTCAGCGTTCGGGTACTCCGGT 1011  
QY 349 LeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrp 368  
DB 1012 TTAGCCATTGACCGCAAAACCTGACACCATCATGTGTTGTTCCCAAGTGTCTTGGTGG 1071  
QY 369 ProAspThrIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrpAsp 388  
DB 1072 CCTGACATGTAGTATTCGTCTCCACCGACCGGGGAAACATGTGTCCCTCTATCGGGAA 1131  
QY 389 TrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrp 408  
DB 1132 CTGACGGTTTCACACCGCGGTACAAAGCAGTACACACCGATTACTCAGGGGTCCTCTGG 1191  
QY 409 LeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTrpMetAsp 428  
DB 1192 CTTGACTTTGGCAATACAGCCAAAGAGCCTGAAGCTAACCCCAAACTCGGGTGGATGACC 1251  
QY 429 GluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThr 448  
DB 1252 CAGTCCTTCGAGATTGACCGCGATTAATCCCGACCGCTTCTTCTATGGCAGCGCGCGGCG 1311  
QY 449 LeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaPro 468  
DB 1312 ATCTATGGCGGACCAACCTCACAACCTGGGATAGGGCAAAAAGTTGATATCACTGTC 1371  
QY 469 MetValLysGlyLeuGluGluThrAlaValAsnAspLeuIleSerProProSerGlyAla 488  
DB 1372 AAGGCCCAAGGATCGAAGAACCCGACGCCCAAGACCTTGGCGCTCCCTCGTGAACATT 1431  
QY 489 ProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaVal 508  
DB 1432 GCCCTATACAGCGCTGCGACATCGGTGGCTTCCACCCCAAAAGATATCTCTCAGGTG 1491  
QY 509 ProSerThr---IlePheThrSerProValPheThrThrGlyThrSerValAspTyrAla 527  
DB 1492 CTTAACAAGTACTACTACAGAACCCACCATGACACCGTGACCTCCATGACTTCGCTCGCC 1551  
QY 528 GluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSerGlnProAsn 547  
DB 1552 GAAAGCAAAACCGCAACTGCTCGTGTGCGAGCAAA--TCTATTCTGGGGAACCC 1605  
QY 548 AspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluPro 567  
DB 1606 ACCTCATGGGTTGGGGTTTCAACTGATGTGCGGAAACGTGGAAACCTGGCGCCACGCCA 1665  
QY 568 GlyGlyValThrThrGlyGlyThrValAlaAlaAspAlaAspGlySerArgPheValTrp 587  
DB 1666 TCAGGGGTTAAAGGCCCGCGCTCCATCACTGTGTGAGCTTAATGCTTCATCCATTGTGTGG 1725  
QY 588 AlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAla 607  
DB 1726 GCCCG-----GAAGGTGCCCGCTCTCGACGCTTCCACCAACAGTGGCTCTCTCGTGCTCA 1779

QY 608 AlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThr 627  
DB 1780 ACGGTGACGGGCTTCTCACAATGCCAGGTTCGATCAGACCGAGTCAACCGTAATACT 1839  
QY 628 PheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGln 647  
DB 1840 CTTTACGGCTCGTTGACGGAAAGTTCTATCACTCCACAAATGGTGGGCTCTCTCAAG 1899  
QY 648 ProValAla---AlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaVal 666  
DB 1900 GCGTCAGCTTTCACGGGTTTCCACACGATGGA-----AACGTTCTGTTCCGTCGCGTG 1953  
QY 667 ProGlyLysGluGlyAspLeuTrpLeuAla-----AlaSerSerGlyLeu 681  
DB 1954 CCGGCGCTCAAGGTCACTCTGGCTTGTGTGGAGTATCGGAAGACCACTATGGGATG 2013  
QY 682 TyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaVal 701  
DB 2014 TGGCGGTCCACTGATGGGGGAAGAACTGGACCAAGGTATCTCGAGTTCAGGAAGCGAT 2073  
QY 702 AsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValValGly 721  
DB 2074 GCGGTGTTTGGTAAAGCAACCAAGTAGTAGCGGATACCCAGTGATTTTCAATCCGCC 2133  
QY 722 ThrIleGlyGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpValLeu 741  
DB 2134 AAGATTGATGAGTTCGAGGAATTTTCGGTCTGACGATGAAGGCAAGACCTTGGAAACGC 2193  
QY 742 IleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAla 761  
DB 2194 ATCAACGACCAACCAACAGTGTATTTGGACTGGCGCATCAATCACCGGTGACCCAGAT 2253  
QY 762 AsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGly 781  
DB 2254 GTCGAAGGCGCGCTACATTGTTACCAACCGTTCGCGCATTTCTCGTGAT----- 2307  
QY 782 GlyAlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeuSer 801  
DB 2307 ----- 2307  
QY 802 ProSerProSerProSerSerProSerProSerProSerProSerSerSerProSer 821  
DB 2308 -----TCTCTCACTCCCGCCCCCGGTCAAGATAGCGAGCGCATCG 2352  
QY 822 SerSerProSerProSerProSerPro---SerProSerProSerArgSerProSerPro 840  
DB 2353 GGAATATACAGGGGCCACACCGACGCTGAAACCCCTGCAGCGATCACAACACTGCGCGG 2412  
QY 841 SerAlaSerProSerProSerSerProSerProSerProSerProSerSerProSerPro 860  
DB 2413 TCCACTGATACGGGCAAACTTCAGAGCCGGGAACAGGAAGTGGATCTGGAAACCGCCT 2472  
QY 861 SerProThrProSerSerSerProValSerGlyValLysValGlnTyrLysAsnAsn 880  
DB 2473 GACTCGGGGAACACCTCTCTCGGAATACCACTTCATGCAAGTTCGCTATTCACGACA 2532  
QY 881 AspSerAlaProGlyAspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySer 900  
DB 2533 GACTGGGCGAGCGG-----TTCACTGGGCGAGCGACCATCACTAACACGCTCCGCC 2583  
QY 901 SerSerValAspLeuSerThrValThrValArgTyr----- 912  
DB 2584 AGCCCAATTAACGGCTGGACGTTGCGGTCACTTACCTTCGCGGCGAGACCATCATGTTCC 2643  
QY 913 ---TyrPheThrArgAsp-----GlyGlySerSerThrLeuValTyrAsnCysAspTrp 929  
DB 2644 GTGTGGTCAGCAACTCAGACTCTGTCTGTCCTCATGTTGTCTCAAGAACTCTGTGTGG 2703  
QY 930 AlaAlaIleGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 949  
DB 2704 -----AATCCACGATACCT 2718

RESULT 5







```

Qy 379 GlyGlyAlaThrTrpThrArgIleTrpAspThrThrSerTyrProAsnArgSerLeuArg 398
Db 1039 TCTGGGACAAATGAGCCCGATCTGGCGTGGCGAGCTATCCGACTGACGACCTATTAC 1098
Qy 399 TyrValLeuAspIleSerAlaGluProTrpLeuThrPheGly--ValGlnProAsnPro 417
Db 1099 TACAGCATCTCAACTCCCAAGACCGGTGGATCAAGAACAACTTTATCGATGTGACGAGC 1158
Qy 418 ProValProSer-----ProLysLeuGlyTrpMetAspGluAlaMetAlaIle 433
Db 1159 GAGTCACCGTCCGATGCTCTCATCAAGCGCTCGCGTGGATGATGCTCTCGAGATT 1218
Qy 434 AspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsn 453
Db 1219 GACCAACCGCAGCAACCACTGGCTCTACGGCACCGGAATGACANATCTTTGGCGGCCAC 1278
Qy 454 AspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeu 473
Db 1279 GATCTCACCACCTGGGACACCGCCACCAATGTGTCAATCCATCACTCGGCAGCGGCATC 1338
Qy 474 GluGluThrAlaValAsnAspLeuIleSerProProSerGlyAlaProLeuIleSerAla 493
Db 1339 GAGGAATCTCGCTCAGGACCTGGCTCTGCACCGCGGAAGCGAGCTATTGGCCGCA 1398
Qy 494 LeuGlyAspLeuGlyGlyPheThrHisAla-----AspValThrAlaValProSer 510
Db 1399 GTCGGAGACGACAAACGGCTTCCACTTTGCCAGCAGAAACGACCTCGGACATCGCCGAG 1458
Qy 511 ThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsn 530
Db 1459 ACGTCTCGGCACACCCACATGGCCACCTCGACGAGCTCGACTACCGCGGAACTCG 1518
Qy 531 ProSerIleValArgAlaGlySerPheAspProProSerSerGlnProAsnAspArgHis 550
Db 1519 GTCAGAGCGTGTCCGCTCGCAACACCGCGCGCAACAG-----1563
Qy 551 ValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyVal 570
Db 1564 GTGGCCATCTCGTCCGCGCGCGCGCGAGCATCGACTACGCGCGCGACACGTC 1623
Qy 571 ThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGly 590
Db 1624 ATGAACGGCGGACCGTGGCTTATTCGGCCGAGCGGACACGATCTCTGTGTCAGCCGC 1683
Qy 591 AspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGln 610
Db 1684 TCGTCCGCG-----GTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCCTCCGCTCG 1734
Qy 611 GlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAla 630
Db 1735 AGCTTCCCGCGCGCGCGCTCATCCCTCGGACAGAAAGACCAACAGCGCTCTTCTACGCC 1794
Qy 631 LeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAla 650
Db 1795 GGCTCCGATACGACCTTTTACGTCAAGGACACCGGACGAGCTTC-----ACG 1845
Qy 651 AlaGlyLeuProSerSerGlyAlaValGlyValMet-----PheHisAlaValProGly 668
Db 1846 CGCGG---CCCAAGCTGGCAGCGGAGCATCCGGATATCGCTGCTCACCCGACC 1902
Qy 669 LysGluGlyAspLeuTrpLeuAlaSerSerGlyLeuTyrHisSerThrAsnGlyGly 688
Db 1903 ACCGGCGGCGCCTGTATGTCTCGACCGAGCTCGGCATATTCGGCTCCACAGACTCGGCG 1962
Qy 689 SerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsnValGlyPheGlyLys 707
Db 1963 ACGACCTTGGCAGATCTCCACCGCCCTGACCAACCTACCATGATCGCCCTGGGTGG 2022
Qy 708 SerAlaProGlySerSerTrpProAlaPheValValGlyThrIleGlyGlyValThr 727
Db 2023 GGCTCA---GGTCTGAATCG---AACCTGTATGCTTCGGACCC-----GGCCCGTCA 2070
728 GlyAla-----TyrArgSerAspCysGlyThrThrTrpValLeuIleAsnAspAsp 745
2071 GGGGCTCGCCTCTAGCCAGTGGAGACAGCGCGCTCTCTGGACGACATCCAGGCGCTCC 2130
746 GlnHisGlnTyrGlyAsnTrpGlyClnAlaIleThrGlyAspHisAlaAsnLeuArgArg 765
2131 CAGGCGTTCGGCTCCATCGACAGCACCAAGGTCCCGCGCAGCGGACGCCCGCGGCAA 2190
766 ValTyrIleGlyThrAsnGlyArgGlyIleValTyr-----GlyAspIleGlyAla 783
2191 GTCTACGTGGGACCAACACCGCGCGCTCTTTTACGCTCAGGAAACCGTCCGCGCGCGC 2250
784 ProSerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerProSer 803
2251 ACGGCGGGGACT-----TCTCTCGTCCAGCAAGCAGC 2283
804 ProSerProSerSerProSerProSerProSerProSerProSerProSerSerSer 823
2284 AGCAGCAGTACTCTTCGCCACGCTCGAGCACACGCTGAGGTGAGCGTTGTATCCACG 2343
824 ProSerProSerProSerProSerProSerProSerProSerProSerProSerProSer 843
2344 ACCCGGGTTCGACGCTGCTGTCGAGGACGACTCGCGCGCGCTCCACCGGCGTCA 2403
843 rProSerPro 846
2404 GGGTTCGCG 2413
RESULT 6
ADH51580
ID ADH51580 standard; cDNA; 2710 BP.
XX
AC ADH51580;
XX
DT 25-MAR-2004 (first entry)
XX
DE Trichoderma reesei eg16 cDNA sequence related to ethanol production.
XX
KW EGV1 protein; endoglucanase; ethanol production; biomass composition;
KW fermentation; cellulohydrolase; detergent production; softening agent;
KW cotton fabric; wood pulp degradation; sugar; gene; ss.
XX
OS Hypocrea jecorina.
XX
FH Key Location/Qualifiers
FT CDS 98..2557
FT /*tag= a
FT /product= "Mature T reesei EGV1 protein"
FT /note= "This sequence contains numerous translation
FT exceptions where the amino acid is given in the protein
FT sequence as J"
XX
PN US2003113732-A1.
XX
PD 19-JUN-2003.
XX
PF 18-DEC-2001; 2001US-00026994.
XX
PR 18-DEC-2001; 2001US-00026994.
XX
PA (DUNN/) DUNN-COLEMAN N.
PA (GOSD/) GOEDEGEBUUR F.
PA (WARD/) WARD M.
PA (YAOJ/) YAO J.
XX
PI Dunn-Coleman N, Goedegebuur F, Ward M, Yao J;
XX
XX WPI; 2004-106460/11.
XX P-PSDB; ADH51581.
XX
XX New substantially purified EGV1 polypeptide with the biological activity
XX of endoglucanase, useful for degrading biomass to ethanol.
XX

```







QY 591 AspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGln 610  
 Db 1724 TCGTCCGC-----GTGACGCTCGCAGTTCCAGGGCAGCTTTGCCTCGCTCTCG 1774  
 QY 611 GlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLeuThrPheTyrAla 630  
 Db 1775 AGCTCCCGCGGGCGCGCTCATCCCTCGGACAGAGACCAACAGCGCTCTTCTACGCC 1834  
 QY 631 LeuSerAsnGlyThrPheTyrArgSerThrAspGlyValThrPheGlnProValAla 650  
 Db 1835 GGCTCCGATCGACCTTTTACGTCAGCAAGACACCGGACGAGCTTC-----ACG 1885  
 QY 651 AlaGlyLeuProSerSerGlyAlaValGlyValMet-----PheHisAlaValProGly 668  
 Db 1886 CGCGGG---CCCAAGCTGGCAGCGCAGGGACGATCCGGATATCGCTCTCACCCGACC 1942  
 QY 669 LysGluGlyAspLeuTrpLeuAlaAsnSerGlyLeuTyrHisSerThrAsnGlyGly 688  
 Db 1943 ACCGCGGCGCAGCTTGATGTCTCGACCGACGTCGCGCATATTCGCTCCACAGACTCGGGC 2002  
 QY 689 SerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsnValGlyPheGlyLys 707  
 Db 2003 ACGACCTTTGGCAAGTCTCCACCGCCCTGACCAACACTACCATGCTCGCTGGGTGTG 2062  
 QY 708 SerAlaProGlySerSerTyrProAlaValPheValGlyThrIleGlyGlyValThr 727  
 Db 2063 GGCTCA---GGCTCAACTGG---AACCTGATGCTTCGGCACC-----GGCCCGTCA 2110  
 QY 728 GlyAla-----TyrArgSerAspAspCysGlyThrTrpValLeuIleAsnAspAsp 745  
 Db 2111 GGGGCTCGCCTCTACCGCAGTCGACAGCGCGGCTCTCTGACGACATCCAGGGCTCC 2170  
 QY 746 GlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArg 765  
 Db 2171 CAGGCTTCGGCTCCATCGACAGCAACCAAGTCGCGCGCAGCGCAGCACCGCGGCAA 2230  
 QY 766 ValTyrIleGlyThrAsnGlyArgGlyIleValTyr-----GlyAspIleGlyGlyAla 783  
 Db 2231 GTCTACGTGGGCACCAACGCGCGGGCGGCTTTTACGTCAGGGAACGTCGCGCGCGCGC 2290  
 QY 784 ProSerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerProSer 803  
 Db 2291 ACGGCGGGGACT-----TCTCTGTCGACCAAGCAGCAGC 2323  
 QY 804 ProSerProSerSerProSerProSerProSerProSerProSerProSerSerSer 823  
 Db 2324 AGCAGCAGTACCTCTTCCGCCAGCTCGAGCACCCAGCTCGAGGTGAGCGTTGTATCCACG 2383  
 QY 824 ProSerProSerProSerProSerProSerProSerProSerProSerProSerAlaSe 843  
 Db 2384 ACCGCGGCTTCACCGGTGACTTCTGTCAGGACAGCTCGCGCGCGCTCCACCGGGGTCA 2443  
 QY 843 rProSerPro 846  
 Db 2444 GGGGTCCGCG 2453  
 RESULT 7  
 ID ADD42060  
 AC ADD42060 standard; DNA; 2849 BP.  
 AC ADD42060;  
 XX  
 DT 15-JAN-2004 (first entry)  
 XX Trichoderma reesei Family 74 xyloglucanase gene, SEQ ID NO:1 #2.  
 DE Trichoderma reesei; Family 74 xyloglucanase; beta-1,4-glycosidic linkage;  
 KW xyloglucan hydrolysis; biomass degradation; cellulose; hemicellulose;  
 KW ethanol production; detergent composition; fabric treatment;  
 KW textile treatment; gene; ds.  
 XX  
 OS Hypocrea jecorina.

Key Location/Qualifiers  
 CDS 59..2743  
 FT /\*tag= b  
 FT /product= "Family 74 xyloglucanase"  
 FT /note= "Contains 3 introns"  
 FT 59..381  
 FT /\*tag= a  
 FT /number= 1  
 FT 382..440  
 FT /\*tag= c  
 FT /number= 1  
 FT 441..1227  
 FT /\*tag= d  
 FT /number= 2  
 FT 1228..1277  
 FT /\*tag= e  
 FT /number= 2  
 FT 1278..2664  
 FT /\*tag= f  
 FT /number= 3  
 FT 2665..2723  
 FT /\*tag= g  
 FT /number= 3  
 FT 2724..2743  
 FT /\*tag= h  
 FT /number= 4  
 WO2003089598-A2.  
 30-OCT-2003.  
 17-APR-2003; 2003WO-US011831.  
 19-APR-2002; 2002US-0373987P.  
 (NOVO ) NOVOZYMES BIOTECH INC.  
 Michael R, Zaretsky E, Haas J;  
 WPI; 2003-845528/78.  
 P-PSDB; ADD42055.  
 New polypeptides having Family 74 xyloglucanase activity, and encoding nucleic acid molecules, useful for degrading cellulose- and hemicellulose -containing biomass to ethanol or as a detergent.  
 Claim 1; Fig 1; 96pp; English.  
 The invention relates to a Family 74 xyloglucanase (ADD42055) from the fungus Trichoderma reesei (Hypocrea jecorina), and nucleic acids encoding it (e.g., ADD42054). The enzyme hydrolyses the beta-1,4-glycosidic linkages in the backbone of xyloglucan to xyloglucan oligosaccharides. The invention also relates polypeptide sequences at least 70% identical to the enzyme, expression vectors and host cells comprising a nucleic acid of the invention, the recombinant production of the enzyme, and mutant enzymes and the nucleic acids encoding them. The xyloglucanase of the invention can be used in the degradation of cellulose- and hemicellulose-containing biomass to produce ethanol. It can also be used in a detergent composition for treating fabric during a machine washing cycle. The present sequence represents the gene encoding Trichoderma reesei Family 74 xyloglucanase. Note: A comparison of this sequence with that shown in the sequence listing (also referred to as SEQ ID NO:1) shows that the present sequence is half the length of the sequence listing SEQ ID NO:1 (ADD42054), as the sequence listing SEQ ID NO:1 contains alternate 60 bp blocks of sense strand and complementary strand sequence.  
 Sequence 2849 BP; 594 A; 882 C; 822 G; 551 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 1.25e-46 Length: 2849  
 Score: 1558.50 Matches: 357  
 Percent Similarity: 54.93% Conservative: 128











Qy 219 TyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTyr 238  
Db 1727 GCGGTGCCACCGCGCGCAATCATGTTGGCGGCGAGCAGCGGGT---GGATCGCCTTT 1783  
Qy 239 ValAlaPheAspIleValSerSerSerLeuGlyGlnAlaSerIleThrIlePheValGly 258  
Db 1784 GTCGTGTTGACGAGCAGCGAGTGGCAACACGGCTCGCAACACCGCGCATCTAGTGGGC 1843  
Qy 259 ValAlaAspProAsnAsnProValPheThrSerArgAspGlyGlyAlaThrTyrGlnAla 278  
Db 1844 GTGTCCACCGAACAGACAGCAGCCTGTATGTGTCCGAAGATGCGCGCGCAGTTGGCACCG 1903  
Qy 279 ValProGlyAlaProThrGlyPheIleProHisIleGlyValPheAspProValAsnHis 298  
Db 1904 GTGCGCGGCAACACCGCGCTCGCGCCGAGCCACATGCGCGCGCGAGCGATGGGCAC 1963  
Qy 299 ValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspVal 318  
Db 1964 ---TGGTATCTGAGCTATGGCGACACGCGCGCGCGGACCTGATGCGGGGAGCCTTG 2020  
Qy 319 TrpIysPheSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAsp 338  
Db 2021 TGGAAATTCACCGCGGACAGCGCGCTGGCGTGAGATCAGCCCGATTCCGCGAG---CCA 2077  
Qy 339 ThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThr 358  
Db 2078 GCCAGTGGCGATGATTCGGCTGGGTGGGTGGGTGGATCGCGCAACATCCGACGGTG 2137  
Qy 359 IleMetValAlaThrGlnIleSerTrpTrpProAspThrIlePheArgSerThrAsp 378  
Db 2138 CTGCTGGCGCACACCTTCGCGCGTGGCGCGCGCGAGAGCTGTATCGGACGCTGGAT 2197  
Qy 379 GlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArg 398  
Db 2198 GTGGCAAGCACTGGACCGCTG---TTGGCCGAT 2230  
Qy 399 TyrValLeuAspIleSerAlaGluProTrpTrpLeuThrPheGlyValGlnProAsnProPro 418  
Db 2231 GCGGTGTTTCGATCACAGCGCGCGCGCTGG---ACCGCACATGCCACCGCGAC--- 2281  
Qy 419 ValProSerProLysLeuGlyTyrMetAspGluAlaMetAlaIleAspProPheAsnSer 438  
Db 2282 ---TGGATGGG---GCGCTGGCGATCGATCCGTTCGACGGC 2317  
Qy 439 AspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrp 458  
Db 2318 AACCATGCGCTGTTCTGACCGGCTACCGCATCTGGGCTCGCGCAATCTG--- 2368  
Qy 459 AspSerGlyGlyGlnIleAlaPro---MetVal 470  
Db 2369 ---CAGATTTCGCGCACCGCAGCGCGCTGCACTGGTGTTCAGGAC 2416  
Qy 471 LysGlyLeuGluGluThrAlaValAsnAspLeuIleSerProProSerGlyAlaProLeu 490  
Db 2417 CGTGGCTTGGAAAGAACCGTGGCTGGACCTGCTCAGCCCGATGGTGGCGCGCATCTG 2476  
Qy 491 IleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSer 510  
Db 2477 CTCAGCGCGCTGGCGATACCGCTTCGCGCATGACGACCTGGACCGCGTG---CAG 2533  
Qy 511 ThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsn 530  
Db 2534 TTGCAGTACCGCGCGCCACCGCTGACCAATGGGAAAGCATCGATGCGCGCGCCAGCGC 2593  
Qy 531 ProSerIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHis 550  
Db 2594 CCGCAGTGGGTGGTGGCAGCGGTACCGTGGCGCGCGCGCGCAACACGAAATCCGCGC 2653  
Qy 550 sValAlaPheSerThrAspGlyGlyValAsnTrpPheGln---GlySerGluProGlyG 569  
Db 2654 C---TGTATTACGCGTGTATGGCGGCAAGCAATGGAGCTCGGTTTGGAGTGGAGCCCGCAGC 2710

Qy 569 yValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPr 589  
Db 2711 CGGCGAGCGCGCGCGAGCATTCATGTTGGTGGCGAGTGGTGGTGGCGACC 2770  
Qy 589 oGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSe 609  
Db 2771 GCGCGCGCGCGCAAT---TGGCGCACGTCGCACTTCGGCGCGCATGTGACAGGTGT 2824  
Qy 609 xGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTy 629  
Db 2825 GAAGCCCTTCGCGAACCCCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2884  
Qy 629 rAla-----LeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGl 647  
Db 2885 TGGCGTGGATGTCGCCAGCGGCGAGCTGTACGAGAGCACCGATGCGCGCGCAGTTCCG 2944  
Qy 647 nProValAlaAlaGlyLeu----- 653  
Db 2945 T-----GCGACCGGTGTACAGGTGGCGAGCCCGCGCGCATGAGCGCACCGCGCGCA 2998  
Qy 654 -----ProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysGluGl 671  
Db 2999 GCTCGTCCGAGCCCGTGGCGCGCGGTGTGTATCTGGCCAGCCCGCGCAAG----- 3053  
Qy 671 yAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTr 691  
Db 3054 -----GGTGTGTGTGGTGGCAGCGGT-----CGCT 3082  
Qy 691 pSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGl 711  
Db 3083 CGAGTGTCTCTCACAGCCGCGAGACAGCTCGTGGGCATCGGACGCGTGGCGGCG 3142  
Qy 711 ySerSerTyrProAlaValPheValValGlyThrIleGlyGlyValThrGlyAlaTyrAr 731  
Db 3143 CGGTGGCGCGCGCTGTATCTGGCGCGCGCTGATCTGGCGCGGTGGTGGTGGTGGT 3202  
Qy 731 sSerAspAspCysGlyThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAs 751  
Db 3203 CTCGACCATGGCGCGCGCGTGGCGCGCATCAAGCATGACGCGCACCGTTCGCGGCG 3262  
Qy 751 nTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAs 771  
Db 3263 C---CCGTACAGCGTGCAGCGGTGATCCGCGCATTCGCGCGCTGTCTTCCGCCACCG 3319  
Qy 771 nGlyArgGlyIleValTyrGlyAspIleGlyAlaProSerGlySerProSerProSe 791  
Db 3320 CGCGCGCGCATTTTCTACGCGCATCCGAGATCGAGCATGTCATCTGCTCCGCGA----- 3374  
Qy 791 rValSerSerAlaSerProSerLeuSerProSerProSerProSerProSerSerPro-- 810  
Db 3375 ---CGTCTGTGTCTTACCGTGTACGCGCGCGCGCTGGCCCTGCGCGCGCGCTG 3430  
Qy 811 -----SerProSerProSerProSerProSerProSerProSerProSerPro-- 827  
Db 3431 ACACGCGCGACCTCGAGATCACCGCATCCCGCT---CCACATTTCCGCGCTGTGTTG 3489  
Qy 828 -----ProSerProSerProSerProSerProSerProSerProSerPro-- 836  
Db 3490 CTGGGTGTGTGCTGGCGAGCGCGCGCGCGCGCAACCGCGCATGCGCTGCGTGTGTTT 3549  
Qy 837 -----SerProSerProSerProSerProSerProSerProSerProSerPr 850  
Db 3550 GCCGATGGCGGTGCTGCAGCGCATGAGGCAATGCCGGTATGGGGTGGGGCGCGCC 3609  
Qy 850 oSerProSerSer-----SerProSerSerProSerProThrPro 864  
Db 3610 GCGCGCGCAATCACGTTGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 3661

RESULT 9  
ADH19117  
ID ADH19117 standard; cDNA; 2646 BP.  
XX  
AC ADH19117;







QY 416 -----AsnProProValProSerPro-----LysLeuGlyTyrMetAspGluAla 430  
 Db 1326 GGATACGGTGGCGCGCAGCGTACCGCGCGCTCACCAGATTGGCTGGTATGATGCGCT 1385  
 QY 431 MetAlaLeuAspProPheAsnSerArgMetLeuTyrGlyThrGlyValAlaThrLeuTyr 450  
 Db 1386 GTGCTTATCGATCGTTCAACCCGAGCAGCTGATGATGACGCGCGCGCGCAGCATCTGG 1445  
 QY 451 AlaThrAsnAspLeuThrLys-----TrpAspSerGlyGlyGlnIleHisLe 466  
 Db 1446 GCGACCGACACGCTCTCCCGTGTGAGAGAGACTGG----- 1481  
 QY 467 AlaPro-----MetValLysGlyLeuGluThrAlaValAsnAspLeuLe 482  
 Db 1482 GCGCGAGCTGATCTCCATCGATCGAGGATATCGAGGAGATGCGATCTGTGCTCGCG 1541  
 QY 483 SerProProSerGlyAlaProLeuLeuSerAlaLeuGlyAspLeuGlyGlyPheThrHis 502  
 Db 1542 TCGCCCAAGAGCGCGCGCGCTCTGCGGCGATCGGTGATGATGCGGATGAGCGAC 1601  
 QY 503 AlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThr 522  
 Db 1602 GAGGACCTCACAAG-----CCGAGAGATGTTGTGCGCCCGCTTCTCCAACCTCGAC 1658  
 QY 523 SerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspPro 542  
 Db 1659 AGCATCGACGCTCGCGGCACTTCCCAACGTTGTGTCGCGCGCGGATCTCTCGGACAC 1718  
 QY 543 SerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrp--- 561  
 Db 1719 GAGTACGACAGCGCGTGGCGCGCGTGGTACGCGACTGACGCGGAGACCGCTGGACC 1778  
 QY 562 ---PheGlnGlySerGluProGlyGlyValThrThr-----GlyGlyThrValAla 577  
 Db 1779 ATCTTCCCTACTGCTCTGCGCATGACGCGGAGCCACTACGAGCGAGCAGATGCA 1838  
 QY 578 AlaSerAlaAspGlySerArgPheValTrpAlaPro-----GlyAspProGlyGlnPro 595  
 Db 1839 GTCGCGGCGGCGGCGGCGAGATCGTGTGTCGACCAAGCTTGACGAGCAGCGCTCGGA 1898  
 QY 596 ValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAla--- 614  
 Db 1899 CCGTGTACTCGACGACTATGCGAAGACGTGTCT-----GTTCCCGCTGGC 1946  
 QY 615 -----AsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyr 629  
 Db 1947 GACCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2006  
 QY 630 AlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProVal 649  
 Db 2007 GCTACCGATGCGCGCAAGTCTCTGCTCGACGCGGCGGAGTGTGATGCGCGGCAAG 2066  
 QY 650 AlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLys 669  
 Db 2067 GCGCGCGGCTGTTGCTCACT-----GCGACATCGCTCATGCTCGCGTGAACCCCTGGTG 2120  
 QY 670 GluGlyAspLeuTrpLeu---AlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGly 688  
 Db 2121 GCGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2180  
 QY 689 SerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsnValGlyPheGlyLys 707  
 Db 2181 GCCTGTTTACGAGGCTAGGTACCGCAACGCGACCGCTGCTGAGCGTGGCGCGCCCAAG 2240  
 QY 708 Ser-----AlaProGlySerSerTyrProAlaValPheValGlyThr--- 722  
 Db 2241 TCCAGTTCGAGCGGCAAGAGGCTAGCGCGCTCGCGCGCTCTTATCTGCGGCGACCGAC 2300  
 QY 723 ---IleGlyGlyValThrGlyAlaTyrArgSerAspCysGlyThrThrTrpValLeu 741  
 Db 2301 AAGCTTGAAGGACATCGGCTGTACCGCTCGACGACACGCGGACCTGGAGCGCGC 2360  
 QY 742 IleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAla 761

Db 2361 GTCAATGACGAGGACCAACTACTCTCGGCG-----CCACCATGATCGAGCGCGCCCAAG 2417  
 QY 762 AsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle 780  
 Db 2418 GTCTACGGCGCGTGTATCTAGGACGAAACGCGCGGATCTGTGTACCGCGACCTT 2474  
 RESULT 10  
 ADH19123  
 ID ADH19123 standard; DNA; 2481 BP.  
 XX  
 AC ADH19123;  
 XX  
 DT 11-MAR-2004 (first entry)  
 XX  
 DE Geotrichum sp M128 xyloglucan oligosacch-degrading enzyme-derived DNA.  
 XX  
 KW xyloglucan oligosaccharide degradation; genetic engineering; ds; gene;  
 XX His tag.  
 OS Synthetic.  
 OS Geotrichum sp. M128.  
 FH Key Location/Qualifiers  
 CDS 1..2481  
 FT /\*tag= a  
 FT /product= "Geotrichum sp M128 xyloglucan oligosaccharide-  
 FT degrading enzyme-derived protein"  
 XX  
 PN EP1350844-A2.  
 XX  
 PD 08-OCT-2003.  
 XX  
 PF 25-MAR-2003; 2003EP-00251866.  
 XX  
 PR 25-MAR-2002; 2002JP-00083433.  
 XX  
 PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.  
 XX  
 PI Yaoi K, Mitsuishi Y;  
 XX  
 DR WPI: 2004-100948/11.  
 DR P-PSDB; ADH19124.  
 XX  
 PT Novel xyloglucan oligosaccharide-degrading enzyme with a different  
 PT degradation mechanism from known enzymes.  
 PS Example 7; SEQ ID NO 17; 39pp; English.  
 XX  
 CC The invention relates to a novel xyloglucan oligosaccharide-degrading  
 CC enzyme with a different degradation mechanism from known enzymes. The  
 CC primary structure and polynucleotide structure of the xyloglucan  
 CC oligosaccharide-degrading enzyme provided by the present invention allows  
 CC high-purity polypeptides having a xyloglucan oligosaccharide-degradation  
 CC activity to be prepared at a low cost through a genetic engineering  
 CC process. The current sequence is that of the Geotrichum sp. M128  
 CC xyloglucan oligosaccharide-degrading enzyme-derived DNA of the invention  
 CC which comprises the wild-type enzyme sequence plus a C-terminal region  
 CC including a Histidine tag originating from vector pET29a(+).  
 XX  
 SQ Sequence 2481 BP; 496 A; 836 C; 748 G; 401 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 2,24e-31 Length: 2481  
 Score: 1124.00 Matches: 298  
 Percent Similarity: 48.91% Conservative: 127  
 Best Local Similarity: 34.29% Mismatches: 306  
 Query Match: 21.89% Indels: 138  
 DB: 12 Gaps: 33  
 US-09-917-376-1 (1-957) x ADH19123 (1-2481)  
 QY 52 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 71



Db	13	TACGAGTTCAAGAAATGTCGGATCGCGCGCGGGGTATACATTACCGGGGATTTGCGCGCAC	72	Db	1009	TGGAAGGACGTCACCCAGCTCTCGTCCCGTCCAACCTCGAAGGTAACTGGGGCCACCCG	1061
Qy	72	GluclyAlaProGlyLeuTyrValArgThrAspIleGlyMetTyrArgTrpAsp	91	Qy	394	AsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpIleuThrPheGlyVal	413
Db	73	CAAAGACCCAGACCTGCTGTACGCGCGACGGACATTTGGCGCGCGTACCGCTGGGAC	132	Db	1069	ACTAAGCGCGCGCGGTAC---AAGGACGGCAGCGCTGTTCCGTGGCTCGACTTCAACAAC	1125
Qy	92	AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr	111	Qy	414	GlnPro-----AsnProProValProSerPro-----LysLeuGly	425
Db	133	GCAGGCACGTCCTCAAGTGGATCCCGCTCAACGACTTTTATCGAGGCGCAGGACATGAACATT	192	Db	1126	GGTCCCCAGTGGGGGGGATACGGTGGCGCGCACCGGTACCCCGGCTCCACCAAGTTTGGC	1185
Qy	112	AsnGlyValValSerIleAlaAAspProIleAsnThrAsnLysValTrpAlaAlaVal	131	Qy	426	TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspMetLeuTyrGlyThr	445
Db	193	ATGGGCACCGAGTCGATCGCGCTGGACCCCAACACCCGACAGGCTGTACTTCGCGCAG	252	Db	1186	TGTTGATGATGAGCTGTGCTTATCGATCGCTTCAACCCCGAGCAGCTGATGATGACGACG	1245
Qy	132	GlyMetTyrThr---AsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAsp	150	Qy	446	GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLys-----TrpAspSerGly	461
Db	253	GGCGGTATGTCGGGACGAGTGG-----GGCGGTCTTATGTGTCGAGAC	300	Db	1246	GGGGCGACCATCTGGGCGACCGACAGCTCTCCGGTTCGAGAAGGACTGG-----	1296
Qy	151	GlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGly	170	Qy	462	GlyGlnIleHisIleAlaPro-----MetValLysGlyLeuGluGluThrAla	477
Db	301	CGCGCGCAGTCGTTTACAATCTACGAGTCGCGCTTCCGATGGCGCCACAGCATGGGA	360	Db	1297	-----GGCGCGAGCTGGTACCTCCAGATCGACGGTATCGAGGAGAAATGCG	1341
Qy	171	ArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGly	190	Qy	478	ValAsnAspLeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeu	497
Db	361	CGCAACAATGGCGAGCGCTCGCTGTCAACCCGTTCAACTCGAACAGAGTCTGGATGGGT	420	Db	1342	ATCTGTCGCTCGCTCGCCCAAGAGCGCGCGCGCTCTCTGTGCGGCATCGGTGACATT	1401
Qy	191	AlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThrTrpSerGlnMet	210	Qy	498	GlyGlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProVal	517
Db	421	ACCGGTACA---GAGGTATCTGAAGAGTTCGACCGCGCCCAAGACTTGGACAAACGTC	477	Db	1402	ACGGCATGAAGCACGACGACCTCACCAAG---CCCCAAGAAGATGTTGTGTGCGCCCCAG	1458
Qy	211	ThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGln	230	Qy	518	PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla	537
Db	478	ACGTCCATCCCGACGCGTTC-----ACCAACGATTCGGATACACG	519	Db	1459	TTTCTCAACTCGACAGCATCGCTCGGGCACTTCTCCCAACGCTTGTGCTCCGCGCC	1518
Qy	231	SerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGln	250	Qy	538	GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly	557
Db	520	TCG-----GTCAATTTTCGACCCC-----GAA	540	Db	1519	GGATCTCGGACACAGTACACAGCGCGTGGCGCGGTGCTGACGGACTGACGGC	1578
Qy	251	AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArg	270	Qy	558	GlyLysAsnTrp-----PheGlnGlySerGluProGlyGlyValThrThr-----	572
Db	541	CGTATGGCACCATCTACGGCAGCGGCGACTGCCCCGACGGC---ATGTACGTACGAC	597	Db	1579	GGAGCGCGTGGACCATCTTCCTACCTGCTCCCTCTGTCATGAACGCGACCATACCAG	1638
Qy	271	AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIle-----	287	Qy	573	GlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro-----Gly	590
Db	598	GACGGCGTCTCGTGGGACCGAGTGGCGGGCAGCGCGTCCAGCTGGCTCAACAGGACC	657	Db	1639	GGCAGCACCATTTGCATGTCACGCGCGCAGCAGATCGTGTGTGTCGACCAAGCTTGAC	1698
Qy	288	-----ProHisLysGlyVal	292	Qy	591	AspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGln	610
Db	658	ACGGCGCGTTCGCGACAAGACCGCGCTCGATCGCGCGCGCCCATGAAGTCGCT	717	Db	1699	GAGGAGCTCGGGACCGTGGTACTCGACGACTATGGCAAGCTGGTCT-----	1749
Qy	293	PheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAsp	312	Qy	611	GlyValProAla-----AsnAlaGlnIleArgSerAspArgValAsn	624
Db	718	CTCACCCCC-----AACTCTCTCTAGCTAGCTTACGCCGACTACCTGTGCTCATGGGC	771	Db	1750	---GTTCCGCTGGCGACCTGAAGCCACAGACTGCCAATGTGCTCTCGACAAAGTCCAG	1806
Qy	313	GlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArgIle---	331	Qy	625	ProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyVal	644
Db	772	GTCAGGTCGGCAAGTCTGGCGCCAGAACCGCACCTCGGGCGCTTGGGACGACATTACT	831	Db	1807	GATGGCACCTTCTACGCTACCGTACCGTGGCGCAAGTTCTTCTCTCGACCGCGCGGAAG	1866
Qy	332	-----SerProValProSerThrAspThrAlaAsnAspTyrPheGly	345	Qy	645	ThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHis	664
Db	832	CCCGCGCTCGGCAACTCGTCGCTGCGCCCGCTACAAACAACAGACGTTTCCCTGGCGCGGA	891	Db	1867	TGTTATGCGCCCAAGGGCGCGGACTTGTACAT-----GGCACATCGCTCATGCTGCC	1920
Qy	346	TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle	365	Qy	665	AlaValProGlyLysGlyLeuTrpLeu---AlaAlaSerSerGlyLeuTyrHis	683
Db	892	TTTTCGGGTCTACGCTGACGCGACCAACCCCAACCGTCTCGTGTATCATCC---CTC	948	Db	1921	GTGAACCCCTGGGTGGCGCGACGCTCTGGTGCCTTCCCGAGGGCGGTCTCTCCAC	1980
Qy	366	SerTrpTrpProAspThrIle-----IlePheArgSerThrAspGlyAlaThr	382	Qy	684	SerThrAsnGlyGlySerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsn	702
Db	949	GACCGGACCCGGACCGCGCTCGACAGCATCTACTCTCAACCGATGCGCGCGGAC	1008	Db	1981	TCGACCGACTTTGGCGCTCTTTC	



```
QY 719 ValValGlyThr-----IleGlyGlyValThrGlyAlaTyrArgSerAspCysGly 736
DB 2101 ATCTGGGCGACCGAACAGCTGGAAAGCGACATCGGCGCTGTACCGCTCGGACGACACGCG 2160
QY 737 ThrThrTrpValLeuIleAsnAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIle 756
DB 2161 AGCACTCGACGCGGTCAATGACGAGGACCACTACTCGGCG---CCACCATGATC 2217
QY 757 ThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleVal 776
DB 2218 GAGGCGGACCCCAAGGTCTACGGGCGGTGTATCTAGGACGACGCGCGGTATCTGTG 2277
QY 777 TyrGlyAspIleGlyGlyAlaProSerGlySerProSerProSerValSerProSerAla 796
DB 2278 TACGCGGACCTTACCACCAAGAAGAGC---AACGAGGAGAAAGTCGACCGCAAAAGTGC 2334
QY 797 Ser-----ProSerLeuSer 801
DB 2335 AACGGCCGAGAGGCGACGCTGTATGTGMAAAGGAGATCTGGGTACCTCGTGCCAC 2394
QY 802 ProSerProSerProSerProSerProSerProSerProSerProSerProSe 821
DB 2395 GCGGTTCAATGGCTATATCGGATCGAATTCGAGCTCGTGCACAGCTTGGCGCGCA 2454
QY 821 rSerSer-ProSerProSerPro 828
DB 2455 CTCGAGCACCAACCACCA 2477

RESULT 11
ADH19119
ID ADH19119 standard; DNA; 2367 BP.
AC ADH19119;
XX 11-MAR-2004 (first entry)
DE Geotrichum sp M128 xyloglucan oligosaccharide-degrading enzyme mat DNA.
KW xyloglucan oligosaccharide degradation; genetic engineering; ds; gene;
KW mature.
OS Geotrichum sp. M128.
XX Key Location/Qualifiers
FH 1. 2367
FT /partial
FT /tag= a
FT /product= "Geotrichum sp M128 xyloglucan oligosaccharide-
FT degrading enzyme mature protein"
FT /notes= "No start or stop codon"
XX EP1350844-A2.
XX 08-OCT-2003.
XX 25-MAR-2003; 2003EP-00251866.
XX 25-MAR-2002; 2002JP-00083433.
XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
XX Yaei K, Mitsuishi Y;
XX WPI; 2004-100948/11.
XX P-PSDB; ADH19120.
XX Novel xyloglucan oligosaccharide-degrading enzyme with a different
XX degradation mechanism from known enzymes.
XX Claim 8; SEQ ID NO 13; 39pp; English.
XX The invention relates to a novel xyloglucan oligosaccharide-degrading
```

```
CC enzyme with a different degradation mechanism from known enzymes. The
CC primary structure and polynucleotide structure of the xyloglucan
CC oligosaccharide-degrading enzyme provided by the present invention allows
CC high-purity polypeptides having a xyloglucan oligosaccharide-degradation
CC activity to be prepared at a low cost through a genetic engineering
CC process. The current sequence is that of the Geotrichum sp. M128
XX xyloglucan oligosaccharide-degrading enzyme mature DNA of the invention.
SQ Sequence 2367 BP; 471 A; 797 C; 719 G; 380 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 3,79e-31 Length: 2367
Score: 1117.00 Matches: 281
Percent Similarity: 50.37% Conservative: 124
Best Local Similarity: 34.95% Mismatches: 279
Query Match: 21.75% Indels: 120
DB: 12 Gaps: 31
```

```
US-09-917-376-1 (1-957) x ADH19119 (1-2367)
QY 52 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 71
DB 10 TACGAGTTCAAGATGTCGCGATCGCGCGGGGTACATACCGGGATGTGCGGCAC 69
QY 72 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArgTrpAsp 91
DB 70 CCAAGACCAAGACCTGTGTACGCGCGACGACATGGCGCGGTACCGCTGGAC 129
QY 92 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr 111
DB 130 GCAGGCACGCTCCAAATGGATCCGCTCAACGACTTTATCGAGCGCGGACATGAACATT 189
QY 112 AsnGlyValIleSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 131
DB 190 ATGGGCACCGAGTCGATCGCTGGACCCCAACACCCGACAGGTGTACCTCGCGCAG 249
QY 132 GlyMetTyrThr---AsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAsp 150
DB 250 GGGCGCTATGTGCGCGCAGGTG-----GCGCGTTCATGTGTCGAGAC 297
QY 151 GlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGly 170
DB 298 CGCGGCGAGTCGTTTCAATCTACGAGTCGCGCTCCCGATGGCGGCAACGACATGGGA 357
QY 171 ArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGly 190
DB 358 CGCAACAATGGCGAGCGCTCGCTGTCAACCCGTTCACTCGAACGAGGTCTGGATGGT 417
QY 191 AlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThrTrpSerGlnMet 210
DB 418 ACGCGTACA---GAGGGTATCTGGAAGAGTTCGAGCCGCCCAAGACTGGACAAACGTC 474
QY 211 ThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGln 230
DB 475 ACGTCCATCCGACGCGTTC-----ACCAACGGTATCGGATACACG 516
QY 231 SerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeuGlyGln 250
DB 517 TCG-----GTCAATTTGACCC-----GAA 537
QY 251 AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArg 270
DB 538 CGTAATGGCACCATCTACGCGAGCGGACTGCCCCCGCAGGGC---ATGATCGTCAGCAC 594
QY 271 AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProGlyIlePheIle----- 287
DB 595 GACGCGGTGTCTCGTGGGAGCCAGTGGCGGCGCAGCGCTCCAGCTGCTCAACAGGACC 654
QY 288 -----ProHisLysGlyVal 292
DB 655 ACGGGCGGTTCGCGGACAAAGACCGCGTCGATCGCGCGCAGCCCATGAAAGTCGT 714
QY 293 PheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAsp 312
```



```
Db 715 CTCACCC---AATCTCTACTGTAAGCCGACTACCTCGTGTCCATGGGCG 768
Qy 313 GlySerSerGlyAspValTyrPheSerValThrSerGlyThrTyrPheArgIle--- 331
Db 769 GTACAGTTTCGGCGAAGTTCGGCGAGAACGCGACCTCGGGCGCTCGGACGACATTACT 828
Qy 332 -----SerProValProSerThrThrThrThrAlaAsnAspTyrPheGly 345
Db 829 CCCCGCTCGGCAACTCTGCTGCTGCCCGCTACAAACACAGAGCTTCCTCGCGGGCGGA 888
Qy 346 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 365
Db 889 TTTTTCGGTCTCAGGCTCGACGCGACCAACCCACCGCTCTCGTCTCATCAC---CTC 945
Qy 366 SerTrpTrpProAspThrIle-----IlePheArgSerThrAspGlyGlyAlaThr 382
Db 946 GACCGCGACCCGCGGACCGCTCGACAGCATCTACCTCTCAACCGATCGCGGCGGAC 1005
Qy 393 TrpThrArgIleTyrAspThrSer-----TyrPro 393
Db 1006 TGAAGGAGCTACCCAGCTCTGCTCCCGCTCCAACTCGAAGTAACTGGGGCCACCCG 1065
Qy 394 AsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTyrPheGlyVal 413
Db 1066 ACTAACGCGCGCGGTAC---AAGACGCGACGCTGTTCCGTGGCTCGACTTCAACAAC 1122
Qy 414 GlnPro-----AsnProValProSerPro-----LysLeuGly 425
Db 1123 GGTCCCAAGTGGGGGGATAGCGTGGCGCGACGCTACCGCGGCTCACCAGTTTGGC 1182
Qy 426 TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr 445
Db 1183 TGGTGGATGAGCGCTGTCTATCGATCCGTTCAACCCGAGCACCTGATGTACGGCAG 1242
Qy 446 GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLys-----TrpAspSerGly 461
Db 1243 GGGGCGACATCTGGCGCCGACGACGCTCTCCGCTGTCGAGAGGACTGG----- 1293
Qy 462 GlyGlnIleHisIleAlaPro-----MetValLysGlyLeuGluGluThrAla 477
Db 1294 -----GCGCGAGCTGTGTACCTCCAGATCGACGGTATCGAGGAGAAATCG 1338
Qy 478 ValAsnAspLeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeu 497
Db 1339 ATCTGTCTCGCTCCGCTCCCAAGAGCGCGCGCGCTCTCTGTGGGATCGGTGACATT 1398
Qy 498 GlyGlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProVal 517
Db 1399 AGCGCATGAGACGACGACTCACCAAG---CCCCAGAGATGTTTGTGGCGGCCAG 1455
Qy 518 PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla 537
Db 1456 TTCTCCAACTCGACAGCATGACGCTCGGGCAACTTCCCAAGTGTGTGTGGCGGCC 1515
Qy 538 GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 557
Db 1516 GGATCTCTCGGGACACGATACGACGCGCTGCGCGCGGTGTGTACGCGACTGACGGC 1575
Qy 558 GlyLysAsnTrp-----PheGlnGlySerGluProGlyGlyValThrThr----- 572
Db 1576 GGAGACGGTGACCATCTTCTTACCTGCGCTCTGTCATGACGCGGACGACCTACCCAG 1635
Qy 573 GlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro-----Gly 590
Db 1636 GGCACACGATTGTCAGTCGACGCGGACGCGGACGACGATCGTGTGTGTGACCAAGCTTAC 1695
Qy 591 AspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGln 610
Db 1696 GAGCAGGCGCTCGGGACCGTGTGTCTCGCACGACTATGCGCAAGACGCTGTCT----- 1746
Qy 611 GlyValProAla-----AsnAlaGlnIleArgSerAspArgValAsn 624
```

```
1747 ---GTTCCGCTGGCGACCTCGAAGCCGACACTGCAATGTCTCTCGGACAAGGTCCAG 1803
Qy 625 ProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyVal 644
Db 1804 GATGGCAGTTTCTACGTTACCGATCCGATGGCGGCAAGTTCTTCTGCTCGACCGCGCGGAAG 1863
Qy 645 ThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHis 664
Db 1864 TCGTATGCGCGCAAGGGCGCGGACTTGTCACT-----GGCACATCGCTCATGCTGCC 1917
Qy 665 AlaValProGlyLysGluGlyAspLeuTrpLeu---AlaAlaSerSerGlyLeuTyrHis 683
Db 1918 GTGAACCCCTGGTGGCGCGGACGCTCTGGGTGCTGTCCGAGGGCGGCTCTCTTCCAC 1977
Qy 684 SerThrAsnGlyGlySerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsn 702
Db 1978 TCGACCGACTTGGCGCTCTGTTACGAGGGTAGGTACCGCAACGCGACCCCTCTGTGAGC 2037
Qy 703 ValGlyPheGlyLysSer-----AlaProGlySerSerTyrProAlaValPhe 718
Db 2038 GTCGCGCCCCCAAGTCCAGTCGCGACGCAAGAGGCTAGCGCGCTCGCGGTCTTC 2097
Qy 719 ValValGlyThr-----IleGlyGlyValThrGlyAlaTyrArgSerAspCysGly 736
Db 2098 ATCTGGGACCGCACAGCCTGGAAGCAGCATCGGCTGTACCGCTCCGACGACACGCGC 2157
Qy 737 ThrThrTrpValLeuIleAsnAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIle 756
Db 2158 AGCAGCTCGACGCGCTCAATGACGAGGACGACAACTACTCGGCG---CCCACCATGATC 2214
Qy 757 ThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleVal 776
Db 2215 GAGCGCGACCCCAAGTCTACGGCGCTGTATCTAGGACGACGCGCGGTATCGTG 2274
Qy 777 TyrGlyAspIle 780
Db 2275 TAGCGCGACCTT 2286

RESULT 12
ADR90295
ID ADR90295 standard; DNA; 2755 BP.
XX
AC ADR90295;
DT 16-DEC-2004 (first entry)
DE Geotrichum sp. M128 xyloglucanendohydrolase precursor DNA.
KW endo-type xyloglucanendohydrolase;
KW xyloglucan-specific endo-b-1,4-glucanase; glucose; xylose; xyloglucan;
KW plant cell differentiation; precursor; ds; gene.
XX
Geotrichum sp. M128.
FH Key Location/Qualifiers
FT CDS 202..2532
/*tag= b
/*product= "Geotrichum sp. M128 xyloglucanendohydrolase
/*precursor protein"
FT sig_peptide 202..261
/*tag= a
/*label= Signal_peptide
FT mat_peptide 262..2529
/*tag= c
/*product= "Geotrichum sp. M128 xyloglucanendohydrolase
mature protein"
XX
JP2004261037-A.
PN 24-SEP-2004.
XX
XX 28-FEB-2003; 2003JP-00053286.
XX
```







```
QY 517 ValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleValArg 536
Db 1693 ACCTTCAGCAGCAGCGCGCTGACCTTCGCGGCGCGCGCAATGCTCGCGCGC 1752
QY 537 AlaGlySerPheAsp-----ProSerSerGlnProAsnAspArgHisValAlaPheSer 554
Db 1753 GTCGGCGCAGCAGCATCTGACCGCGCGCTCGCGGGTGCACGACGAGGGCGGTACACG 1812
QY 555 ThrAspGlyGlyLysAsnTrp-----PheGlnGlySerGluProGly---GlyValThr 571
Db 1813 ACCAACACGCGCGCAGCAGCTGACGCTCTTTTCAGACCTGTGTCCCGCAGCTCGAGTCCGT 1872
QY 572 ThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGlyAsp 591
Db 1873 AACGGCGCAGCATCGCGCTCGCGCGCAGCGCAGACGTTTGTGTGTCGCCAGCAAG 1932
QY 592 ---ProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGln 610
Db 1933 GCGGACGCAAGGGCGGTACACATCGAGCGACTACGCGCAAGACGTGGACTGCGCGCAGC 1992
QY 611 GlyValProAlaAsnAlaGln---IleArgSerAspArgValAsnProLysThrPheTyr 629
Db 1993 GGCCTGAGCAAGCAGACGACAGGTATCGCGCGCAGCGCGCTGCGCGCCACACACATCTTAC 2052
QY 630 AlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProVal 649
Db 2053 GTGTACGTGAGGGCGCATCTTCTGCTCGACCGCATGCGCGCAGAGCTACACCAAGAAG 2112
QY 650 AlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGly--- 668
Db 2113 GGCAACGCTCCGCTGCTGTGACG-----TACACGGGCGACCGCGTCAAG 2160
QY 669 -----LysGluGlyAspLeuTrpLeuAlaSerSer---GlyLeuTyrHisSer 684
Db 2161 AGCAACTTGCCTGCGCGCGAGCTGTGGTCTCCGTCAGGCGCGTGGCATCTACCAACAGC 2220
QY 685 ThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGly 704
Db 2221 ACCGACTTTGGCAACACATTCACCGCGCTTGGCGGCTCGGGCTCAGTCTCAAT----- 2274
QY 705 PheGlyLysSerAlaProGlySerSerTyrProAlaValPheValValGlyThr----- 722
Db 2275 -----CCCGCTGTGTTCAGCATCGGCGCGCGCAG 2304
QY 723 -----IleGlyVal-----Thr 727
Db 2305 ACGCCAACGCCAGCAGCGCTCTCTCGGGCATCCGCTCGGCGTGCAGCGCGCAG 2364
QY 728 GlyAlaTyrArgSerAspCysGlyThrThrTrpValLeuIleAsnAspGlnHis 747
Db 2365 GGCCTGTATATGTCGACCGCAACGCGCGGTTGTGGACGCGCCTCAACGACGCGCGCAG 2424
QY 748 GlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgValTyr 767
Db 2425 AACTATGTGGC---GCCACCGTATCAGCGCGGATCCCGCATCTACGCGCGGTCTTAC 2481
QY 768 IleGlyThrAsnGlyArgGlyVal 776
Db 2482 ATTGCGATGACGCGCGGCGCATCATC 2508
RESULT 13
ID ADR90297
AC ADR90297 standard; DNA; 2268 BP.
XX
AC ADR90297;
XX
DT 16-DEC-2004 (first entry)
DE Geotrichum sp. M128 xyloglucanendohydrolase mature DNA.
XX
KW endo-type xyloglucanendohydrolase;
KW xyloglucan-specific endo-b-1,4-glucanase; glucose; xylose; xyloglucan;
```

```
KW plant cell differentiation; mature; ds; gene.
XX Geotrichum sp. M128.
OS
XX
FH Key Location/Qualifiers
FT CDS 1..2268
FT /tag= a
FT /product= "Geotrichum sp. M128 xyloglucanendohydrolase
FT mature protein"
FT /note= "No start or stop codon"
PN JP2004261037-A.
XX
XX 24-SEP-2004.
XX
XX 28-FEB-2003; 2003JP-00053286.
XX
XX 28-FEB-2003; 2003JP-00053286.
XX (DOKU-) DOKURITSU GYOSEI HOJIN SANGYO GIJUTSU SO.
XX
XX WPI; 2004-665466/65.
XX P-PSDB; ADR90298.
XX
XX Novel xyloglucan hydrolase derived from Geotrichum genus, having ability
XX to hydrolyze beta-1,4-glycoside linkage of glucose residue in
XX xyloglucan, useful for elucidating structure or function of xyloglucan.
XX
XX Claim 7; SEQ ID NO 14; 86pp; Japanese.
XX
XX The invention relates to a novel xyloglucanendohydrolase (xyloglucan-
XX specific endo-b-1,4-glucanase) enzyme derived from the Geotrichum genus
XX and having the ability to hydrolyse the beta-1,4-glycosidic linkage of a
XX glucose residue, but not that of a xylose residue contained in a
XX xyloglucan. The polypeptide of the invention may be useful for
XX elucidating the structure or function of a xyloglucan having an important
XX role in the differentiation of a plant cell. The current sequence is that
XX of the Geotrichum sp. M128 xyloglucanendohydrolase mature DNA of the
XX invention.
XX
XX Sequence 2268 BP; 413 A; 795 C; 717 G; 343 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 2,666-26 Length: 2268
Score: 979.00 Matches: 264
Percent Similarity: 48.43% Conservative: 122
Best Local Similarity: 33.12% Mismatches: 279
Query Match: 19.07% Indels: 132
DB: Gaps: 33
US-09-917-376-1 (1-957) x ADR90297 (1-2268)
QY 57 ValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGly 76
Db 22 GTGACGATCTCGGTGGCGGCTTCATCTCGGGCCTCGTGGCACACCGCGCAGAGGAC 81
QY 77 IleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaValArg 96
Db 82 CTGATCTACGCGCAGCAGACATTTGGCGGACGTTACAGGTGGAAACGCGCGCAAGTGGAG 141
QY 97 TrpIleProLeuLeuAspTrpVal-----GlyTrpAsnAsnTrpGlyTyr 111
Db 142 TGGGAGCCCATCACGGACTTTTATCATCAACAACGCGCTCGCGGCAACCGCGCAACCTG 201
QY 112 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 131
Db 202 CTCGGCAGGAGAGCATTTGCGCTCGACCCCAACACCGCGCTGTACCTCGCGCAA 261
QY 132 GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGln 151
Db 262 GCGCACTATGTG---CAGTGGGACCGG---TGGGCGGCGCTTCTCTCTCGGACGACGCG 315
```



152 GlyAlaThrTrpGlnIleThrProLeuProPheLeuGlyGlyAsnMetProGlyArg 171  
172 GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 191  
376 AATGCGGCGAGCGCTCGCGGTCAACCGCACCTGGACGAGCTGTGGTTGGCTCG 435  
192 ProSerGlyLeuGlyLeuTyrArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 211  
436 CGCAGC---CAGGGCTGTGGCGCTCCACTGACCGCGCTCAGACGTGGTCGCGATGAAC 492  
212 AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrGlyTyrGlnSer 231  
493 CAGCTCCCGCAGCTCAGCTGAC---GGAATC----- 522  
232 AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeuGlyGlnAla 251  
523 -----GGCATCTCGGTTCATCTCGACCCCAAGACGTCGGCAGCATACGTC 573  
252 SerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArgAsp 271  
574 GCTTCGACGCGCTCGGCGGCTGTGG-----GTCACGTGG-----GAC 612  
272 GlyGlyAlaThrTrpGlnAlaValProGlyAlaProThr----- 284  
613 GCGCGCGCACTGCTCCAGGTGGCGCGCAGCTTACGCACTGTGTACAGTGTGGAACAAG 672  
285 -----GlyPheIleProHisLysGly 291  
673 AGCATTTGTCGTGCTCGGCACCGCATCCAGAGCTCGGGCCCTCCCGATCAAGATC 732  
292 ValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyr 311  
733 GCACGTGGCAAG---AAGCGCAGGCTGTACATTACTCTCGACGCGACCGGCGCCCTGG 789  
312 AspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArgIle 331  
790 GCGCTCTCTACGCGGAGTGTGGTCTACGATCCACCAACGCGCACTGGAAGCACATC 849  
332 SerProValProSerThrAspThrAlaAsnAspTyr----- 343  
850 ACC-----CCCTCGCGCGAGCGGCGCAACAGTACCTCGCCCGCCAGCAACAAAG 903  
344 -----PheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMet 360  
904 GTCGTTCCGGAGCTGGAACGATCTCGGTGCT-----AAGCGCGACACCGTCGTC 957  
361 ValAlaThrGlnIleSerTrpProAspThrIlePheArgSerThrAspGlyGly 380  
958 GTGTCGACGCTCGACGCAACCGCGAAGACTCA---GTGTACCTCTCGCGCGACGCGCGC 1014  
381 AlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrVal 400  
1015 AACTGTGGAAGACCTCGGCAAGCTGCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1074  
401 LeuAspIleSerAlaGlu-----ProTrpLeuThrPheGlyVal 413  
1075 AAGAGTTCGCGCAAGCTGCGCAACGCGCATCCCGCTCCCTGGCTCTCGTTCAGAAC 1134  
414 GlnProAsnProProValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIle 433  
1135 CGCGGCTCGGTATTGTC-----GGATTGCGGTGTGGTCTCGCGCGCATCTCTCTC 1185  
434 AspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsn 453  
1186 GACCCCTTC---TCGACCGCTCTCTGTACGCGCACAGCGCGCTGTGATCTGGCGCATG 1242  
454 AspLeuThrLysTrpAspSerGlyGlnIleHisIleAlaProMet----- 469  
1243 CCGCTGTGCGCGCGCGCTCG-----AACCAGCGCGCGAGCTGGTACATCAAC 1290  
470 ValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProProSerGly-----Ala 488

1291 ACAGAGGCGCATCGAGGAGACGCGCATCTGTGTCCTCAAGTCGCGGCTCGGGCGCGCG 1350  
489 ProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaVal 508  
1351 CACCTCTTCAGCGGATGTACGACCTCGCGGCGATCGGCATCATGACTTC---TCGCTC 1407  
509 ProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGlu 528  
1408 CCGAGCCCATGTACAGCAAGCGACCTTCAGCACGACGACGCGCTCGACTTTGCGGCG 1467  
529 LeuAsnProSerIleIleValArgAlaGlySerPheAsp-----ProSerSerGlnPro 546  
1468 CCGCGCGCCATGTGCTCGCGCGCTCGCGCGCAACGACCATCTCGACGCGCGCTCGCG 1527  
547 AsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrp-----PheGlnGly 564  
1528 GGGTGCACGACGCGCGGTACAGCACCAACAGCGGCGAGCTGGAGCTCTTTCAGACC 1587  
565 SerGluProGly---GlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySer 583  
1588 TGTGTCCCGAGCTCGAGTGTGTACGCGGCGCACGATCGCGTTCGCGCGGCGGCAAG 1647  
584 ArgPheValTrpAlaProGlyAsp---ProGlyGlnProValValTyrAlaValGlyPhe 602  
1648 ACGTTTGTGTGTCGCCCGCAGCGGCGCAAGCGCGCGCGCGCTACACATCGAGCGACTAC 1707  
603 GlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGln---IleArgSerAsp 621  
1708 GGCACAGCTGAGCTCGCGCGCGCGCTGAGCAACGACGACGAGATATCGCGCGCGAC 1767  
622 ArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAsp 641  
1768 CCGGTGACGCGCAACACATTTCTACGTGTAGTTCGAGCGGCGACTTCTTCGTCTCAGCCGAT 1827  
642 GlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyVal 661  
1828 GCGCGCAAGAGCTACACCAAGAGGCGCAACGCGCTCCCGTCTCGTGGACG----- 1878  
662 MetPheHisAlaValProGly-----LysGluGlyAspLeuTrpLeuAlaAla 677  
1879 ---TACACGCGGACGCGCGTACAGCAACTTTCGTGCGGCGGAGCTGTGGGTCTCGCTC 1935  
678 SerSer---GlyLeuTyrHisSerThrAsnGlySerSerTrpSerAlaIleThrGly 696  
1936 AAGGCGTGTGGCATCTACCAACGACCGACTTTGGCAACACATTCACCGCGCTTCGCGGC 1995  
697 ValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAla 716  
1996 TCGGGCTCCAGTCTCAAT-----CCCGCT 2019  
717 ValPheValValGlyThr-----IleGlyGly 725  
2020 GTGTTCAGCATCGCGCGCGCGCAGACGCCCAACGCGCGAGCGCTCTCTCTCGGGGC 2079  
726 Val-----ThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrp 739  
2080 ATCCCGTTCGCGCTCGCAGCCCGAGGCGCTGTATGTGCGACCGCAACGCGGGGTGG 2139  
740 ValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAsp 759  
2140 ACGCGCTCAACGACGACGCGGCAACTATGGTGGC---GCCACCGTGTATCAGCGCGCAT 2196  
760 HisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleVal 776  
2197 CCGCGCATCTACGCGCGCTCTACATTGGCATGAACGCGCGGGGCATCATC 2247  
RESULT 14  
ADR90301  
ID ADR90301 standard; DNA; 2274 BP.  
XX  
AC ADR90301;  
XX



DT 16-DEC-2004 (first entry)  
 XX Geotrichum sp. M128 xyloglucanendohydrolase mature DNA + Met.  
 XX  
 XX endo-type xyloglucanendohydrolase;  
 KW xyloglucan-specific endo-b-1,4-glucanase; glucose; xylose; xyloglucan;  
 KW plant cell differentiation; mature; ds; gene.  
 XX  
 XX Geotrichum sp. M128.  
 OS Synthetic.  
 XX  
 XX Location/Qualifiers  
 FH 1. 2274  
 FT CDS /tag= a  
 FT /product= "Geotrichum sp. M128 xyloglucanendohydrolase  
 FT mature protein + Met"  
 XX  
 PN JP2004261037-A.  
 XX  
 XX 24-SEP-2004.  
 XX  
 XX 28-FEB-2003; 2003JP-00053286.  
 XX  
 XX 28-FEB-2003; 2003JP-00053286.  
 XX  
 XX (DOKU-) DOKURITSU GYOSEI HOJIN SANGYO GIJUTSU SO.  
 XX  
 XX WPI; 2004-665466/65.  
 DR P-PSDB; ADR90302.  
 XX  
 XX Novel xyloglucan hydrolase derived from Geotrichum genus, having ability  
 PT to hydrolyze beta-1, 4-glycoside linkage of glucose residue in  
 PT xyloglucan, useful for elucidating structure or function of xyloglucan.  
 XX  
 XX Example 6; SEQ ID NO 18; 86pp; Japanese.  
 XX  
 XX The invention relates to a novel xyloglucanendohydrolase (xyloglucan-  
 CC specific endo-b-1,4-glucanase) enzyme derived from the Geotrichum genus  
 CC and having the ability to hydrolyse the beta-1,4-glycosidic linkage of a  
 CC glucose residue, but not that of a xylose residue contained in a  
 CC xyloglucan. The polypeptide of the invention may be useful for  
 CC elucidating the structure or function of a xyloglucan having an important  
 CC role in the differentiation of a plant cell. The current sequence is that  
 CC of the Geotrichum sp. M128 xyloglucanendohydrolase mature DNA of the  
 CC invention which has DNA encoding an additional Met residue at the N-  
 CC terminal.  
 XX  
 SQ Sequence 2274 BP; 415 A; 795 C; 719 G; 345 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 2,668-26 Length: 2274  
 Score: 979.00 Matches: 264  
 Percent Similarity: 48.43% Conservative: 122  
 Best Local Similarity: 33.12% Mismatches: 279  
 Query Match: 19.07% Indels: 132  
 DB: 13 Gaps: 33

US-09-917-376-1 (1-957) x ADR90301 (1-2274)

QY 57 ValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGly 76  
 DB 25 GTGAGATCTCGGTGGGGGCTTCATCTCGGGCTCGTGGGACACCCCGAGGAGGAC 84  
 QY 77 IleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTyrAspAlaAlaAsnGlyArg 96  
 DB 85 CTGATCTAGCAGCGCACAGACATTGGCGGCACGTACAGGTGGAACGCGGCAAGTGGGAG 144  
 QY 97 TrpIleProLeuLeuAspTrpVal-----GlyTrpAsnAsnTrpGlyTyr 111  
 DB 145 TGGGAGCCCATCAGGACTTTATCATCAACAACGCGCTCGCGGGCAACGCGCCACCTG 204  
 QY 112 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaVal 131  
 DB 1189 GACCCCTTC---TCGACCGCTCTCTGTACGGCAGGCGCTGTGATCTGGCGGACTGAC 1245

DB 205 CTCGGCACCGAGCATTGCGCTGCACCCCCCAACCCAGACCGCTGTACCTCGCGCAA 264  
 QY 132 GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGln 151  
 DB 265 GCGGACTATGTG---CAGTGGGACCGG---TGGGGCGCTTCTCTCGGACGACCGC 318  
 QY 152 GlyAlaThrTrpGlnIleThrProLeuProPheLeuLeuGlyGlyAsnMetProGlyArg 171  
 DB 319 GGCAGAGCGTTCAAGCAGCAGTACCGCTCGCCCGTCCCATCGGCGCAACACATGGGCGC 378  
 QY 172 GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 191  
 DB 379 AATGGCGGCGAGCGCTCGCGGTCAACCGGCATCGGACCGAGGTGTGGTTGGCTCG 438  
 QY 192 ProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 211  
 DB 439 CGCAGC---CAGGGCTGTGGCGCTCCACTGACCGCGCTCAGACGTGTGTCGCATGAAC 495  
 QY 212 AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSer 231  
 DB 496 CAGCTCCCGCACAGCTCGAGGTAC-----GGAATC----- 525  
 QY 232 AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGlnAla 251  
 DB 526 -----GGCATCATCTCGGTTCATCTTCGACCCCAAGAACGTCGCGCATACGTC 576  
 QY 252 SerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArgAsp 271  
 DB 577 GCTTCGACCGCTCGGGGCTGTGG-----GTCACTGG-----GAC 615  
 QY 272 GlyGlyAlaThrTrpGlnAlaValProGlyAlaProThr----- 284  
 DB 616 GCGCGGCCCACTGGTCCCGAGGTGGCGCGCCAGCTACGAGTGTGTCAGACTGGACAAAG 675  
 QY 285 -----GlyPheIleProHisLysGly 291  
 DB 676 AGCATTTGCTGCTGCTCGGCGCACCGCATCCAGAGCTCGGGGCGCTCCCGATCAAGATC 735  
 QY 292 ValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyr 311  
 DB 736 GCACCTGGGCGAAG---AACGGCAGGCTGTATACCTACTCGGACGACCGGCGCCCTGG 792  
 QY 312 AspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyTrpTrpThrArgIle 331  
 DB 793 GCGCTCTCTAGCGGAGGTCTGGTCTAGATCCCAACCAACGCGCACTGGAAGCACATC 852  
 QY 332 SerProValProSerThrAspThrAlaAsnAspTyr----- 343  
 DB 853 ACC-----CCCTCGCGCGAGGCGGCGCAACACAGTACCTGCCCCCACTGGCAACAAAAAG 906  
 QY 344 -----PheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMet 360  
 DB 907 GTCGTTCCCGGAGGCTGGAAACGGTATCTCGGTGGT-----AACCGGACACCGTCGTC 960  
 QY 361 ValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGly 380  
 DB 961 GGTGACCTCTGACGCAACCGCGAAGACTCA---GTGTACCTCTCGCGGACGCGCGC 1017  
 QY 381 AlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrVal 400  
 DB 1018 AACTCGTGAAGGACCTCGGCAAGCTGACACCGCCCGCGGCGGCGCAACTCGCAG 1077  
 QY 401 LeuAspIleSerAlaGlu-----ProTrpLeuThrPheGlyVal 413  
 DB 1078 AAGGAGTCGAGCGCAAGCTGGCGCAACGCGCACTTCGCTCCCTCGGTCTCTGTTCGAAC 1137  
 QY 414 GlnProAsnProProValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIle 433  
 DB 1138 CGCGGCTCGGGTATTGTC-----GGATTGCGGTGGTGGCTCGCGGCACTCTCTCTC 1188  
 QY 434 AspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsn 453  
 DB 1189 GACCCCTTC---TCGACCGCTCTCTGTACGGCAGGCGCTGTGATCTGGCGGACTGAC 1245



QY	454	AspLeuThrLysTrpAspSerGlyGlnIleHisIleAlaProMet-----	469
DB	1246	GCGGTGTCGCGCGGACTCG-----ACACAGGCGCGGAGCTGTGTACATCAAC	1293
QY	470	VallysGlyLeuGluGluThrAlaValAsnAspLeuIleSerProProSerGly---Ala	488
DB	1294	ACAGAGGCATCGAGAGCGGATCTGTCTCAAGTCGCGCCTGCGGCGCCGCG	1353
QY	489	ProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaVal	508
DB	1354	CACCTCTTCAGCGCATCTAGACCTCGCGCATCGCCATGATGACTTC---TCCGTC	1410
QY	509	ProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGlu	528
DB	1411	CCGAGCCCATGTATACAGAACCGACCTTCAGCAGCAGCGGCTCGACTTCGCGGC	1470
QY	529	LeuAsnProSerIleIleValArgAlaGlySerPheAsp-----ProSerSerGlnPro	546
DB	1471	CGCGGCGCAATGTCTCGCGCGCTCGGCGCGCAACGACCATCTCTGACGCGGCGTCGG	1530
QY	547	AsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrp-----PheGlnGly	564
DB	1531	GGGTGACGCGCGGCGGTACAGCAACAGCGCGCACAGCTGACGCTCTTTCAGACC	1590
QY	565	SerGluProGly---GlyValThrThrGlyThrValAlaAlaSerAlaAspGlySer	583
DB	1591	TGTGTCCCGACGCTCGAGTCGTAAACGCGCGCACGATCGCGGTGCGCGCGCAAG	1650
QY	584	ArgPheValTrpAlaProGlyAsp---ProGlyGlnProValValTyrAlaValGlyPhe	602
DB	1651	ACGTTGTGTGTCTGCCCGCAGCAAGCGGAGCGGCGGCGGCGGCGGCGGCGGCTAC	1710
QY	603	GlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGln---IleArgSerAsp	621
DB	1711	GCGACAGCTGTGACTGCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	1770
QY	622	ArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAsp	641
DB	1771	CGCGTGCAGGCCAACACATCTTACGTGTACGTGCGGCGGCGGCGGCGGCGGCGG	1830
QY	642	GlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyVal	661
DB	1831	GCGGCGACAGGTACACCAAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	1881
QY	662	MetPheHisAlaValProGly-----LysGluGlyAspLeuTrpLeuAlaAla	677
DB	1882	----TACCGGCGCGCGCGCTCAGCAGCAACTTGTGCGGCGGCGGCGGCTGTGCGT	1938
QY	678	SerSer---GlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGly	696
DB	1939	AAGGGGTTGGCATCTTACACAGCAGCGGCTTGTGGCAACACATTCACCGGCTTGG	1998
QY	697	ValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAla	716
DB	1999	TCGGGCTCCAGTCTCAAT-----CCGCT 2022	
QY	717	ValPheValValGlyThr-----IleGlyGly 725	
DB	2023	GTGTTTCAGCATCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	2082
QY	726	Val-----ThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrp	739
DB	2083	ATCCCGTCGCGGTCGCGAGCGCGGCGGCTGTATATCTCGACCGACACGCGGCTGT	2142
QY	740	ValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAsp	759
DB	2143	ACGCGCTCAACGACGCGGCGCAACATATGTGGC---GCCACCGGTGATCAGCGCGAT	2199
QY	760	HisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleVal 776	
DB	2200	CCCCGCACTTACGCGCGCGGTCTTACATTTGGCATGACGCGCGGCGGCGATCATC	2250

RESULT 15  
ADD42054  
ID ADD42054 standard; DNA; 5698 BP.  
XX  
AC ADD42054;  
XX  
15-JAN-2004 (first entry)  
XX  
Trichoderma reesei Family 74 xyloglucanase-encoding DNA, SEQ ID NO:1 #1.  
XX  
Trichoderma reesei; Family 74 xyloglucanase; beta-1,4-glycosidic linkage;  
KW xyloglucan hydrolysis; biomass degradation; cellulose; hemicellulose;  
KW ethanol production; detergent composition; fabric treatment;  
KW textile treatment; ds.  
XX  
OS Hypocrea jecorina.  
XX  
PN WO2003089598-A2.  
XX  
PD 30-OCT-2003.  
XX  
PF 17-APR-2003; 2003WO-US011831.  
XX  
PR 19-APR-2002; 2002US-0373987P.  
XX  
PA (NOVO ) NOVOZYMES BIOTECH INC.  
XX  
PI Michael R, Zaretsky B, Haas J;  
XX  
WPI; 2003-845528/78.  
DR P-PSDB; ADD42055.  
XX  
New polypeptides having Family 74 xyloglucanase activity, and encoding  
PT nucleic acid molecules, useful for degrading cellulose- and hemicellulose  
PT -containing biomass to ethanol or as a detergent.  
XX  
Claim 1; SEQ ID NO 1; 96pp; English.  
XX  
The invention relates to a Family 74 xyloglucanase (ADD42055) from the  
CC fungus Trichoderma reesei (Hypocrea jecorina), and nucleic acids encoding  
CC it (e.g., ADD42054). The enzyme hydrolyses the beta-1,4-glycosidic  
CC linkages in the backbone of xyloglucan to xyloglucan oligosaccharides.  
CC The invention also relates polypeptide sequences at least 70% identical  
CC to the enzyme, expression vectors and host cells comprising a nucleic  
CC acid of the invention, the recombinant production of the enzyme, and  
CC mutant enzymes and the nucleic acids encoding them. The xyloglucanase of  
CC the invention can be used in the degradation of cellulose- and  
CC hemicellulose-containing biomass to produce ethanol. It can also be used  
CC in a detergent composition for treating fabric during a machine washing  
CC cycle. The present sequence represents a Trichoderma reesei Family 74  
CC xyloglucanase-encoding DNA. Note: the present sequence (referred to as  
CC SEQ ID NO:1) is given as shown in the sequence listing. A comparison of  
CC this sequence with that shown in figure 1 (also referred to as SEQ ID  
CC NO:1) shows that the present sequence is twice as long as the figure 1  
CC sequence (ADD42060) and contains alternate 60 bp blocks of sense strand  
CC and complementary strand sequence.  
XX  
SQ Sequence 5698 BP; 1145 A; 1704 C; 1704 G; 1145 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 1,36e-24 Length: 5698  
Score: 940.00 Matches: 374  
Percent Similarity: 30.11% Conservative: 136  
Best Local Similarity: 22.08% Mismatches: 315  
Query Match: 18.31% Indels: 871  
DB: 10 Gaps: 59  
US-09-917-376-1 (1-957) x ADD42054 (1-5698)  
QY 25 SerPheAlaValAlaAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAla 44  
DB 128 TCTCGATCTCTGCTGTCTGTCTGCGGCGGCGGCTC-----ATCCCTGCC 169



QY 45 HisAlaAlaThrThrGln----- 50  
|||||  
Db 170 CATGCTGCTTCTTCCAGAGAGCTCAGGAACGGGAACAGGACCCCGCGCAGTAGGGACGG 229  
QY 51 -----ProTyrThrTrpSerAsnValAlaIle---GlyGlyGlyGlyPheValAspGly 67  
|||||  
Db 230 GTACGACGGATTTCATGGAAGACGTCAAGCTCGCGCGCGCGCGCTTCGTCGCCCGCG 289  
QY 68 IleValPheAsn-----GluGlyAlaPro----- 75  
|||||  
Db 290 ATCATCTTCCAAAGTACCTTCTTCAGATTCCAGCGCGCGCGCGCGCGCGGAACGAGGGCGCG 349  
QY 76 -----GlyIleLeuTyrValArgThrAspIleGlyGlyMet 87  
|||||  
Db 350 TAGTAGAAGGTTCCCAAGACAAAGGCGTGCATATGCACGAACAGATATTGGCGGGCTG 409  
QY 88 TyrArgTrp----- 90  
|||||  
Db 410 TACCGCTCAAAAGGTTCTGTTTCCGCATCGTATACGTGTTGTCATATACCGCGCGAC 469  
QY 91 -----AspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpVal----- 104  
|||||  
Db 470 ATGGCGAGATTTCGCCGACGACTCATGGACCGCGCTCACGGATGGGATTGCTGATAATGCC 529  
QY 105 GlyTrp-----AsnAsnTrpGly----- 110  
|||||  
Db 530 GGCTGGCAACAGCGGCTGCTGAGTACCTGCGCGGCGAGTCCCTAACGACTATTACGG 589  
QY 111 -----TyrAsnGlyValSerIleAlaAlaAspProIleAsnThrAsnVal 127  
|||||  
Db 590 CCGACCGTGTCTGGGGATCGAGCTGTGGCTTGTATCCGCAGGACGATCAAAAGGTG 649  
QY 128 TrpAlaAlaVal----- 131  
|||||  
Db 650 TATGCGCAGTGAACCCGCTAGTCGCAACAGCGCAACTAGGCGTCTGCTAGTTTTCAC 709  
QY 132 -----GlyMetTyrThrAsnSerTrpAsp----- 139  
|||||  
Db 710 ATACGGGTCTCGGCATGTATACGAACAGCTGGTC-TGTGATGTCTCAGATCTAGACC 768  
QY 139 ----- 139  
Db 769 TATGATTGGACGGCGGTACATATGCTTGTGCAACGACACTACAGAGAGTCTAGATCTGG 828  
QY 140 -----ProAsnAspGlyAlaIleLeuArg 147  
|||||  
Db 829 ATACTAACTCGCTGACATTTGCCCATATATAGGATCCGAGTAATGGAGCCATCATTCGC 888  
QY 148 SerSerAspGln----- 151  
|||||  
Db 889 TCGTCAGACCGGACTGTAAACGGGTATATATCCCTTAGGCTCATTTACCTCGGTAGTAAGCG 948  
QY 152 -----GlyValaThrTrpGlnIleThrProLeuProPheLeuLeuGlyGlyAsn 167  
|||||  
Db 949 AGCAGTCTGGGGGCGGCGCAACGTGTCTTCCAACTTGCCCTTCAAAGTCGGGGGTAAAC 1008  
QY 168 MetProGlyArg----- 171  
|||||  
Db 1009 ATGCAGGACCGCGGTTGCAACAGGAAGTGTGTTGAACGGGAAGTTTCAGCGCCCATTTG 1068  
QY 172 -----GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeu 187  
|||||  
Db 1069 TACGCTCTCGGGAGCGGAGAGCGTCTGGTGTGTCATCGATCCGGCCAACTCCAACATCATC 1128  
QY 188 TyrPheGlyAlaPro----- 192  
|||||  
Db 1129 TACTTTGTGCTCTCTCGGCGCTCTCCGACCGGACAGTAGGCGCGGTTGAGGTTGTAGTAG 1188  
QY 193 -----SerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrp 207  
|||||  
Db 1189 ATGAACACGACGCTCAGGAACAGCGCTCTGGAAGTCTACGGACGGCGGCTGACCTTT 1248  
QY 208 SerGlnMet----- 210  
|||||

Db 1249 TCCAAAGTCTCGCGGAGTCTTTTCCGGGAGACCTTCAGATGCTGCGCGCAGCTGGAAA 1308  
QY 211 -----ThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 227  
|||||  
Db 1309 AGGTTCCAGAGCTCGTTTCCAGCAACTGGAGCGTATCATCCAGACCCGAGTGATTCCAAC 1368  
QY 228 GlyTyrGln----- 230  
|||||  
Db 1369 GGCTCAACAGCAGCAAGTGGCGTTGACCTGTCATGTAGGGTCTGGGCTCAGTAAGGTTG 1428  
QY 231 -----SerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSer 247  
|||||  
Db 1429 CCGATGTTGTCGACAGCAAGGACTCATGTGGTTACGTTTCGACTCAACACGACGACG 1488  
QY 248 LeuGlyGlnAla----- 251  
|||||  
Db 1489 ACCGGGGAGCCCTGTTCTCTGAGTACACCAATGCAAGCTGAGTTGGTCTGTCGTGC 1548  
QY 252 -----SerLysThrIlePheValGlyValAlaAspProAsnAsnProValPhe 267  
|||||  
Db 1549 TGGCCCCCTCGGACGCTCGTATCTTTGTGGACGGCTGATAACATCACTGCTTCAGTC 1608  
QY 268 Trp----- 268  
|||||  
Db 1609 TATGTGACACGTGCAGACATAGAAACAAACCGTCCGACTATTGTAGTGACGAAGTCAG 1668  
QY 269 -----SerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 286  
|||||  
Db 1669 ATACACTCGTGAATGCGCGCTCCACGTGAGTGTGTACCGGGGCGACGAGGAATAC 1728  
QY 287 IleProHis----- 289  
|||||  
Db 1729 TTTCTCTCAAGTTACGCGGAGTGCACCTCAGACATGCGCCCTCGTCCCTTTATG 1788  
QY 290 LysGlyValPhe-----AspProValAsnHisValLeuTyrIleAlaThrSerAsn 306  
|||||  
Db 1789 AAAGAGGTGTCGCGAACTGCGACGACGAGAGAGCCCTTGTATCTGACCTATTCCGAT 1848  
QY 307 ThrGlyGlyPro----- 310  
|||||  
Db 1849 GGCACAGGCGCGCTTTGACGTCGTCCTCTTCCGGAACATAGACTGGATAAGGCTA 1908  
QY 311 -----TyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGly 326  
|||||  
Db 1909 CCGTGTCCGCGCTATGATGGCACACTTGGCTCAGTGTGAGGTACGACATTGACGGGGGA 1968  
QY 327 ThrTrpThrArgIleSerProValProSerThrAspThr----- 339  
|||||  
Db 1969 ACTTGGAAAGACATACTACCGTGTGNAACGAGTCAACCTCCATGCTGTAAACGTCGCCCT 2028  
QY 340 -----AlaAsnAspTyrPheGlyTyrSerGlyLeu 349  
|||||  
Db 2029 TGAACCTTTCTGATCACCCCTGCTCTCGGATCAGATCTATCTTGGCTTTGGCGGCTT 2088  
QY 350 ThrIleAsp----- 352  
|||||  
Db 2089 GGCCTCGATTGTAGTGGGGACAGAGACCTAGTCTAGATATGAAACCGAACCCGCGGAA 2148  
QY 353 -----ArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpPro 369  
|||||  
Db 2149 CCGGAGCTAAACAAAGCCAGAACCCCTTGTGTGTCTTTTGAACCTTTGGTGCCCA 2208  
QY 370 AspThr----- 371  
|||||  
Db 2209 GATGCTCAGCTGTTTTCGTCCTTGGGAACAACACGAAGAACTTGAGAACCCACCGCT 2268  
QY 372 -----IleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrpAspTrp 389  
|||||  
Db 2269 CTACGAGTCGACTTTCGTCGACCGACTCTGGGACAAACATGAGCCGATCTGGCGCTGG 2328  
QY 390 ThrSerTyrProAsnArgSer---LeuArgTyrValLeuAspIleSerAlaGluPro--- 407  
|||||



Db 2329 GCGAGCTATCGAAAGCCAGCTGGCTGAGACC-CTGTTGTACCTCGGCTAGACCCGCGAC 2387  
Qy 407 ----- 407  
Db 2388 CCGCTCGATAGGCACTGAGACCTATTACTACAGCATCTCAGTGAGTCACCTTAACGATC 2447  
Qy 408 -----TrrLeuThrPhe-----GlyValGlnProAsn----- 416  
Db 2448 CGATGCGAATGGCTGACTCTGGATAATGATCTGTAGAGTCACCTCAGTGAGAAATTGCTAG 2507  
Qy 417 -----ProProValPro----- 420  
Db 2508 GCTAGCTTACCGACTGACTCTGCTTCAAGACTCCCAAGCACCCTGGATCAAGAACAC 2567  
Qy 420 ----- 420  
Db 2568 TTTATCGATGTGATGACTGAGACGAAGTTCTGAGGGTTTCGTGGCACCCTAGTCTTCTGTTG 2627  
Qy 421 -----SerPro-----LysLeuGlyTrrMet 427  
Db 2628 AAATAGCTTACACTCGAGCGAGTCAACCGTCCGATGCTCATCAAGCGCCTCGGCTGGATG 2687  
Qy 428 AspGluAlaMetAla----- 432  
Db 2688 ATTGAGTCTCTCGGCTCGCTCAGTGGCAGGCTACAGAGTAGTTCCGCGAGCCGACCTTAC 2747  
Qy 433 -----IleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAla 447  
Db 2748 TAACTCAGAGAGCAGATTGACCCACCGACGACCAACCACTGGCTCTCTACCGCACCAGGAATG 2807  
Qy 448 ThrLeuTyrAlaThr----- 452  
Db 2808 ACAATCTTTGGCGTTAACTGGTTGGCTGCTGTTGGTGACCGAGATCCCGTGGCTTTAC 2867  
Qy 453 -----AsnAspLeuThrLysTrrAspSerGlyGlyGlnIleHisIleAla 467  
Db 2868 TGTAGAAACCGCGCCACGATCTCAACACTGGGACACCGGCCACATGTGTCAATCCNA 2927  
Qy 468 ProMetValLysGly----- 472  
Db 2928 TCACCTGGCAGACCGCGTGTAGTGGTTGACCCCTGTGCGGGTGTACACAGTTAGTT 2987  
Qy 473 -----LeuGluGluThrAlaValAsnAspLeuIleSerProSerGly 487  
Db 2988 AGTGACCGCTCGCGCATCGAGGAATTCTCCGTCAGAGACCTGGCCTCTGCACCCCGCGGA 3047  
Qy 488 AlaProLeuIleSer----- 492  
Db 3048 AGCGAGCTATTGGCGTAGCTCCTTAAGAGCGCAGGTCCTGGACCGGAGACGTGGCGCGCT 3107  
Qy 493 -----AlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAla 507  
Db 3108 TCGCTCGATAACCCCGCAGTCGAGACGACCAACGCGCTTCACCTTTGCCACGAGAAACGAC 3167  
Qy 508 Val----- 508  
Db 3168 CTCGGGACATCGCGCGTCTGCTGCTGTTGGCGAAGTGAAACGGTCTTTTGTGCTG 3227  
Qy 509 -----ProSerThrIlePheThrSerProValPheThrThrGlyThrSerVal 524  
Db 3228 GAGCCCTGTAGCGCGCAGCGCTCTGGGCAACGCCACATGGGCCACCTCGACGAGCGTC 3287  
Qy 525 AspTyrAla----- 527  
Db 3288 GACTACGCGCGGAGGCTGTGCAGACCCGTTGCGGGGTGTACCCGGTGGAGCTGCTCGCAG 3347  
Qy 528 -----GluLeuAsnProSerIleValArgAlaGlySer----- 539  
Db 3348 CTGATGCGGCCCTACTCTCGTCAAGAGCGTCGTCGCGTCCGCAACACCCCGCGCAGCAA 3407  
Qy 540 -----PheAspProSerSerGlnProAsnAspArgHisValAla 552  
Db 3408 CAGGTGGCCATCTTGAGCCAGTTCTCGCAGCAGCGCGCGCTCTCTGGAGCCGTTGTGGCGCGCTGCGTT 3467

Qy 553 Phe-----SerThrAspGlyGlyLysAsnTrrPheGlnGlySerGluProGlyGly 569  
Db 3468 GTCCACCGGTAGACGCTCCGACGCGCGCGAGCTGGAGCATCGACTACGCGCGCAGACG 3527  
Qy 570 ValThrThrGlyGly----- 574  
Db 3528 TCCATGAACGGCGCAGGCTGCCGCGCGCTGCACCTCGTAGCTGATGTGCCCGCTGTGC 3587  
Qy 575 -----ThrValAlaAlaSerAlaAspGlySerArgPheValTrrAlaPro 589  
Db 3588 AGTACTTTCCGCGCAGCGTGGCTTATTCGCGCGCAGCGCACGATCTCTGTGTCGACC 3647  
Qy 590 GlyAspProGly----- 593  
Db 3648 GCCTCGTCGCGCGCTGCCACCGGATAAGCCGGCTGCCCTGTGTAGGAGACCACTGG 3707  
Qy 594 -----GlnProValValTyrAlaValGlyPheGlyAsnSerTrrAlaAlaSerGlnGly 611  
Db 3708 CGGACGCGCGCTCGCAGCGCTCG-----CAGTTCCAGGCGAGCTTTGCCCTCCGTCGAGC 3764  
Qy 612 ValProAlaAsnAlaGln----- 617  
Db 3765 CTGCCCGCGCGCGCGACGCTCGCGAGCGTCAAGSTCCCTCGAAACGCGAGCAGAGCTCG 3824  
Qy 618 -----IleArgSerAspArgValAsnProLysThrPheTyrAlaLeu 631  
Db 3825 GACGGCGCGCGCGCTCATCGCTCGGACAGAACACCAACAGCGTCTTCTACGCGCGC 3884  
Qy 632 SerAsnGlyThrPhe----- 636  
Db 3885 TCCGATCGACCTTTTAGTAGCGGACCTGTTCTTCTGTTGTCGCAAGATGCGGCGC 3944  
Qy 637 -----TyrArgSerThrAspGlyGlyValThrPheGlnProValAlaAla 651  
Db 3945 AGGCCTAGCTGCAAAACGTCAGCAAGGACACCGCGCAGCGCTTC-----ACGCGC 3995  
Qy 652 GlyLeuProSerSerGlyAlaValGlyValMet----- 662  
Db 3996 GGG-----CCCAAGCTGGCGCAGCGAGGATGCACTGTTCTGTGGCGCTCGTGGAGTGC 4052  
Qy 663 -----PheHisAlaValProGlyLys 669  
Db 4053 GCGCGCGGGTTGACCCGCTCGCTCCCTCGATCGGGATATCGCTGCTACCCGACCCACC 4112  
Qy 670 GluGlyAspLeuTrrPheAlaAlaSerSer----- 679  
Db 4113 GCGGGCAGCTTGTATGTCGACCGGCTAGGCCCTATAGCGACGAGTGGGTGTGTG 4172  
Qy 680 -----GlyLeuTyrHisSerThrAsnGlyGlySer 689  
Db 4173 CGCCCGTGCAACATACAGAGCTGGCTCGCGCATATTCGCTCCACAGACTCGGGCAGC 4232  
Qy 690 Ser----- 690  
Db 4233 ACCTTTGGCAAAGTCTCCACCGCCCTGAAGCCGTATGAAGCGAGGTGTCTGAGCCGCTGC 4292  
Qy 691 -----TrrSerAlaIleThrGlyValSerSerAlaValAsnValGly 704  
Db 4293 TGGAAACCGGTTTCAGAGGTGGCGGACTCCACACCTTACCAGATCGCCTCGGGTGTGGC 4352  
Qy 705 PheGlyLysSer-----AlaPro 710  
Db 4353 TCAGGCTCGAACTGGAACCTGTATGCTGTGTGTGATGTTCTTAGCGGAGCCCAACACCCG 4412  
Qy 711 GlySerSerTyrProAlaValPheValValGlyThrIleGlyGlyValThrGlyAla--- 729  
Db 4413 AGTCCGAGCTTGACCTTGGACATACAGGATCGGACCC-----GGCCGCTCAGGGGCTCGC 4466  
Qy 730 ---TyrArgSerAspAspCysGlyThrTrrPheVal----- 740  
Db 4467 CTCTACGCCAGTGGAGACAGCGCGCGCTCTCTGGAGCCGTTGGCGCGGCGAGTCCCGAGCG 4526



```
QY 741 -----LeuIleAsnAspAspGlnHisGlnTyrGly--- 750
Db 4527 GAGATGCGGTCACTCTGTGCGCGGAGGAGCTCGGACATCCAGGGCTCCCGAGGGCTTC 4586
QY 750 ----- 750
Db 4587 GGCTCCATCGACAGCACCAAGGTCGCCGCGCAGCGCCTGTAGGTCCCGAGGGTCCCGAAG 4646
QY 751 -----AsnTyrGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgVal 766
Db 4647 CGAGGTAGTGTCTGTGTGTTCCAGCGCGCTCGCGCAGCACCGCCGGG-----CAAGTC 4700
QY 767 TyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAsp----- 779
Db 4701 TACGTGGGCACCAACGGCGCGGGCGCTCTTTACGCTCAGGCGTCTGCGGGCCCGTTCAG 4760
QY 780 -----IleGlyGlyAlaPro 784
Db 4761 ATGCACCGGTGTGTCGCGCCCGCGAGAAATGCGAGTCCGAACCGTCGCGCGCGGC--- 4817
QY 785 SerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerProSerPro 804
Db 4818 ACGGGCGGGACTTCTCTCGTCCAGCAAGCAGCAGCAGCAGTACTTGGCAGCGCCGCCG 4877
QY 805 SerPro-----SerSerSerProSerProSerProSerPro 816
Db 4878 TGCCCGCCCTGAAGAGCAGCTGGTTCGTCTCGTCTGTCTCATCTCTTCCGCCAGCTCG 4937
QY 817 SerSerSerProSerSerSerPro-SerProSerPro----- 828
Db 4938 AGCACACGCTGAGGTCGAGCGTTGTATCCAGACCGCGGCTTGGAGAAGCGGTCGAGC 4997
QY 829 -----SerProSerProSer----- 833
Db 4998 TCGTGTGCGACTCCAGCTCGCAACATAGTGTCTGGGCCCGAACGACGCTGACTTCGTG 5057
QY 834 -----ProSerArgSerProSerProSerAlaSerProSerProSerProSerProSerPr 852
Db 5058 AGGACCAAGCTCGGCCCGCGTCCACAGGGGTGCGCGGCTGCCACTGAAGCAGC 5117
QY 852 oSerSerSerProSerSerProSerProSerProSerProSer 865
Db 5118 TCCTGGTCGAGCGCGCGCCAGGGGTGCCAGTCCCCAGC 5157
```

Search completed: October 5, 2005, 09:19:17  
Job time : 1592.27 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model  
Run on: October 5, 2005, 07:43:21 ; Search time 9202.32 Seconds  
(without alignments)  
3958.515 Million cell updates/sec

Title: US-09-917-376-1  
Perfect score: 5135  
Sequence: 1 MDRSENRLTWRRRLVSL.....RASFGSNPATPTADTYLQX 957

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Pgapop 6.0 , Pgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues  
Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cpn2.1/USPTO spoel/US09917376/runat\_04102005\_164328\_26911/app\_query.fasta\_1.2524  
-DB=EST -QFMT=fast -SURFIX=p2n.rst -MINMATCH=0.1 -LOOPECL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09917376 @CGN 1.1 12009 @runat\_04102005\_164328\_26911 -NCPU=6 -ICPU=3  
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN TIMEOUT=30 -THRAD=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_hic:\*  
4: gb\_est3:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_gsel1:\*  
9: gb\_gsel2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	631.5	12.3	704	CF872104	CF872104 trico29xo
2	631.5	12.3	761	CB902332	CB902332 trico29xo
3	603.5	11.8	747	CF876916	CF876916 trico074xd
4	603.5	11.8	814	CB905388	CB905388 trico074xd
5	415.5	8.1	707	CF880713	CF880713 trico082xn
6	415.5	8.1	782	CB907625	CB907625 trico082xn
7	401.5	7.8	693	CF882065	CF882065 trico29xo
8	362	7.0	748	CE342004	CE342004 tigr-gss-
9	361.5	7.0	751	CF688882	CF688882 trico16xm

10	361.5	7.0	803	6	CB898982	CB898982 trico16xm
11	360	7.0	791	9	CNS028QT	AL186302 Tetraodon
12	358	7.0	285	7	CF706250	CF706250 CCAGP33TF
13	358	7.0	509	7	CF707273	CF707273 CCADL40TF
14	358	7.0	741	7	CF710894	CF710894 CCADL94TO
15	356.5	6.9	1066	3	CR704524	CR704524 Tetraodon
c 16	355	6.9	522	9	CE190185	CE190185 tigr-gss-
c 17	353	6.9	493	8	CC060605	CC060605 CSU-K33r.
18	351	6.8	877	9	CR071250	CR071250 Forward s
c 19	350	6.8	694	8	CC119324	CC119324 NDL.70K21
c 20	347	6.8	259	8	AZ374273	AZ374273 IM0136G21
c 21	347	6.8	587	6	C79507	C79507 C79507 Mous
c 22	340.5	6.6	708	9	AG459371	AG459371 Mus muscu
23	340.5	6.6	783	9	AG468194	AG468194 Mus muscu
24	338	6.6	738	7	CF866388	CF866388 trico06xe
25	338	6.6	794	6	CB896354	CB896354 trico06xe
c 26	333	6.5	703	9	CE176982	CE176982 tigr-gss-
c 27	329.5	6.4	476	6	C79500	C79500 C79500 Mous
c 28	329	6.4	888	7	CV069323	CV069323 WPAEHUX15
29	328	6.4	562	8	AZ855343	AZ855343 2M0159A22
30	325	6.3	807	7	CN783154	CN783154 ai.L003.4
31	322	6.3	1270	9	AG448439	AG448439 Mus muscu
32	314.5	6.1	546	2	BF072664	BF072664 NCSM3H7T3
c 33	314	6.1	619	8	BZ192350	BZ192350 CH230-248
c 34	314	6.1	959	9	CNS022KT	AL221078 Tetraodon
35	311	6.1	726	7	CO549464	CO549464 LYEST7094
36	310	6.0	744	8	BZ277113	BZ277113 CH230-450
c 37	307	6.0	538	9	CC847893	CC847893 NDL.4416
c 38	306.5	6.0	389	5	BU639045	BU639045 mgcw011XD
c 39	306	6.0	961	9	CNS030KT	AL256070 Tetraodon
c 40	301.5	5.9	464	8	AZ928926	AZ928926 479.d1f19
c 41	298.5	5.8	440	8	AZ312209	AZ312209 IM0027L19
c 42	297	5.8	929	7	CF885920	CF885920 trico085xf
c 43	292	5.7	393	8	AZ332252	AZ332252 IM0060H12
c 44	291	5.7	761	9	CNS041RC	AL270561 Tetraodon
c 45	290	5.6	775	9	AG603507	AG603507 Mus muscu

ALIGNMENTS

RESULT 1  
CF872104  
LOCUS trico29xo|3.b1 T.reesei mycelial culture, Version 6 October 2003  
DEFINITION Hypocrea jecorina cdna clone trico29xo|3, mRNA sequence.  
ACCESSION CF872104  
VERSION CF872104.1 GI:38126786  
KEYWORDS EST.  
SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)  
ORGANISM Hypocrea jecorina  
REFERENCE 1 (bases 1 to 704)  
AUTHORS Diener S.E., Dunn-Coleman, N., Foreman, P., Houfek, T.D.,  
Teunissen, P.J.M., van Solingen, P., Dankmeyer, L., Mitchell, T.K.,  
Ward, M. and Dean, R.A.  
TITLE Characterization of the protein processing and secretion pathways  
in a comprehensive set of expressed sequence tags from Trichoderma  
reesei  
JOURNAL FEMS Microbiol. Lett. 230 (2), 275-282 (2004)  
COMMENT Contact: Ralph A. Dean  
Fungal Genomics Laboratory  
North Carolina State University  
Campus Box 7251, Raleigh, NC 27695, USA  
Tel: 919-513-0020  
Fax: 919-513-0024  
Email: ralph.dean@ncsu.edu  
Seq primer: Lr-F1 primer:  
Location/Qualifiers  
1. .704  
/organism="Hypocrea jecorina"  
/mol\_type="mRNA"  
/strain="QM6a"



```
/db_xref="taxon:51453"
/clone="tric029xol3"
/dev_stage="mycelia"
/clone_lib="T.reesei mycelial culture, Version 6 October
2003"
note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial
culture grown from 24 hrs to 6 days with varying Carbon
and Nitrogen sources and concentrations."

ORIGIN

Alignment Scores:
Pred. No.: 1.63e-28 Length: 704
Score: 631.50 Matches: 127
Percent Similarity: 71.55% Conservative: 39
Best Local Similarity: 54.74% Mismatches: 49
Query Match: 12.30% Indels: 19
DB: 7 Gaps: 6

US-09-917-376-1 (1-957) x CP872104 (1-704)
QY 25 SerPheAlaValAlaAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAla 44
DB 47 TCTGAGTCCTTGGCTTGTCTGGGGCCGTC-----ATCCCTGCC 88
QY 45 HisAlaAlaThrThrGlnProTyrThrTrpSerAsnValAlaIle---GlyGlyGlyGly 63
DB 89 CATGCTGCC-----TTTTCATGGAAGAACGTCAAGCTCGGGCGGGCGGGC 136
QY 64 PheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAsp 83
DB 137 TTCGTCCCGGCATCATCTTCCATCCCAAGACAAAAGCGTAGCATATGCAGAACAGAT 196
QY 84 IleGlyGlyMetTyrArg---TrpAspAlaIleAsnGlyArgTrpIlePro-----Leu 100
DB 197 ATTCGGCGGCTGTACCGCTCTTCTCCGCGACTCATGACCGCC-GTCAAGGATGGGATT 255
QY 101 LeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAsp 120
DB 256 GCTGATAATGCGGCTGGCAACATGG-----GGCATCGACGCTGTTCGGCTTGAT 306
QY 121 ProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspPro 140
DB 307 CGCAGGACGATCAAAAGGTGTATCCGAGTCCGATGTATACGAACAGCTGGGATCCG 366
QY 141 AsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeu 160
DB 367 AGTAATGAGGCATCATTCGCTCGTCAGACCGCGCGCAACGTGCTTCCACCAACTTG 426
QY 161 ProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAsp 180
DB 427 CCTTTCAAAGTCGGGGGTAACATGCAGACGCGGAGCGGAGCGCTGGCTGTGAT 486
QY 181 ProAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSer 200
DB 487 CGGCGCAACTCCACATCATCTACTTTGGTGTCTGCTCAGGAACGGCTCTGGAAGTCT 546
QY 201 ThrAspSerGlyAlaThrTrpSerGlnMetThrAsnProAspValGlyThrTrpIle 220
DB 547 ACGGACGCGCGGCTGACCTTTTCCAAAGTCTCGTGTTCACGGCACTGGGACGTACATC 606
QY 221 AlaAsnProThrAspThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAla 240
DB 607 CCAGACCGCGGTGATTTCAACGGCTTACCACAGCGATCAGCAAGGACTCATGTGGGTACG 666
QY 241 PheAspLysSerSerSerLeuLeuGlyGlnAlaSer 252
DB 667 TTGACTC-ACCAGCAGCAGCAGCCGNGAGCCACG 701

RESULT 2
CB902332 761 bp mRNA linear EST 02-JUL-2003
LOCUS tric029xol3 T.reesei mycelial culture, Version 3 april Hypocrea
DEFINITION jecorina cDNA clone tric029xol3, mRNA sequence.

ACCESSION CB902332
VERSION CB902332.1 GI:30116990
EST.
KEYWORDS Hypocrea jecorina (anamorph: Trichoderma reesei)
SOURCE Hypocrea jecorina
ORGANISM Hypocrea jecorina
REFERENCE 1 (bases 1 to 761)
AUTHORS Foreman,P.K., Brown,D.E., Dankmeyer,L., Dean,R., Diener,S.,
Dunn-Coleman,N.S., Goedegebuur,F., Houfek,T.D., England,G.J.,
Kelley,A.S., Meerman,H.J., Mitchell,T., Mitchinson,C.,
Olivares,H.A., Teunissen,P.J., Yao,J. and Ward,M.
TITLE Transcriptional regulation of biomass-degrading enzymes in the
filamentous fungus Trichoderma reesei
JOURNAL J. Biol. Chem. 278 (34), 31988-31997 (2003)
MEDLINE 22803314
PUBMED 12788920
COMMENT Contact: Pamela K. Foreman
Genencor Intl.
925 Page Mill Road, Palo Alto, CA 94304, USA
Tel: (650) 846-7635
Fax: (650) 621-7817
Email: Pforeman@genencor.com
Seq primer: LT-F1 primer.
FEATURES
source
1..761
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM6a"
/db_xref="taxon:51453"
/clone="tric029xol3"
/dev_stage="mycelia"
/clone_lib="T.reesei mycelial culture, Version 3 april"
/note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial
culture grown from 24 hrs to 6 days with varying Carbon
and Nitrogen sources and concentrations."

ORIGIN

Alignment Scores:
Pred. No.: 1.8e-28 Length: 761
Score: 631.50 Matches: 127
Percent Similarity: 71.55% Conservative: 39
Best Local Similarity: 54.74% Mismatches: 49
Query Match: 12.30% Indels: 19
DB: 6 Gaps: 6

US-09-917-376-1 (1-957) x CB902332 (1-761)
QY 25 SerPheAlaValAlaAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAla 44
DB 104 TCTGAGTCCTTGGCTTGTCTGGGGCCGTC-----ATCCCTGCC 145
QY 45 HisAlaAlaThrThrGlnProTyrThrTrpSerAsnValAlaIle---GlyGlyGlyGly 63
DB 146 CATGCTGCC-----TTTTCATGGAAGAACGTCAAGCTCGGGCGGGCGGGC 193
QY 64 PheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAsp 83
DB 194 TTCGTCCCGGCATCATCTTCCATCCCAAGACAAAAGCGTAGCATATGCAGAACAGAT 253
QY 84 IleGlyGlyMetTyrArg---TrpAspAlaAlaAsnGlyArgTrpIlePro-----Leu 100
DB 254 ATTGCGGGCTGTATCCGCTCTTCTCCGCGACTCATGGAACCGCC-GTCAAGGATGGATT 312
QY 101 LeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAsp 120
DB 313 GCTGATAATGCGGCTGGCACAACATGG-----GGCATCGACGCTGTTCGCTTGAT 363
QY 121 ProfileAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspPro 140
DB 364 CGCAGGACGATCAAAAGGTGTATCCGAGTCCGAGTGTATATACGAACAGCTGGGATCCG 423
QY 141 AsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeu 160
```



```

Db 424 AGTAATGGAGCATATTTCGTCGTACAGCGCGGCAACGTGTGCTCTCACCACATTTG 483
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Qy 161 ProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAsp 180
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 484 CCCTTCAAGTCCGGGGTAAATGATCCAGAGCGCGAGCGGAGAGCGTCTGGCTGTGAT 543
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Qy 181 ProAsnAsnAspAsnIleLeuTy-PheGlyAlaProSerGlyLysGlyLeuTrpArgSer 200
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 544 CGGCGCAACTCCACATCATCTACTTTGGTGTCTGCTCAGGAACCGGCTCTGGAAAGTCT 603
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Qy 201 ThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrlle 220
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 604 ACGAGCGCGGGTGGTACCTTTTCCAGGTCTCGTTCACGGCACTGGGACGTATATC 663
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Qy 221 AlaAsnProThrAspThrThrGlyTyrglnSerAspIleGlnGlyValValTrpValala 240
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 664 CCAGACCGGAGTGATTCCCAACGGCTACCAAGCGATCAGCAAGGACTCATGTGGGTAGC 723
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Qy 241 PheAspLysSerSerSerSerLeuGlyGlnAlaSer 252
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 724 TTGACTC-ACCAGCAGCAGCAGCGGNGGAGCCACG 758
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::

```

RESULT 3  
CF876916 747 bp mRNA linear EST 31-OCT-2003  
LOCUS  
DEFINITION  
tric074xd02.b1 T.reesei mycelial culture, Version 6 October 2003  
Hypocrea jecorina cDNA clone tric074xd02, mRNA sequence.

ACCESSION  
CF876916  
VERSION  
CF876916.1 GI:38131598  
KEYWORDS  
EST.  
SOURCE  
Hypocrea jecorina (anamorph: Trichoderma reesei)  
ORGANISM  
Hypocrea jecorina  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.

REFERENCE  
AUTHORS  
Diener,S.E., Dunn-Coleman,N., Foreman,P., Houfek,T.D.,  
Teunissen,P.J.M., van Solingen,P., Dankmeyer,L., Mitchell,T.K.,  
Ward,M. and Dean,R.A.  
TITLE  
Characterization of the protein processing and secretion pathways  
in a comprehensive set of expressed sequence tags from Trichoderma  
reesei

JOURNAL  
FEMS Microbiol. Lett. 230 (2), 275-282 (2004)

COMMENT  
Contact: Ralph A. Dean  
Fungal Genomics Laboratory  
North Carolina State University  
Campus Box 7251, Raleigh, NC 27695, USA  
Tel: 919-513-0020  
Fax: 919-513-0024  
Email: ralph.dean@ncsu.edu  
Seq primer: Tr-F1 primer.

FEATURES  
source

1..747  
Location/Qualifiers

/organism="Hypocrea jecorina"  
/mol\_type="mRNA"  
/strain="QM6a"  
/db\_xref="taxon:51453"  
/clone="tric074xd02"  
/dev\_stage="mycelia"  
/clone\_lib="T.reesei mycelial culture, Version 6 October  
2003"  
/note="Vector: pREP3Y; Site\_1: Not I/Sal I; Mycelial  
culture grown from 24 hrs to 6 days with varying Carbon  
and Nitrogen sources and concentrations."

ORIGIN

Alignment Scores:  
Pred. No.: 8.65e-27 Length: 747  
Score: 603.50 Matches: 121  
Percent Similarity: 65.43% Conservative: 38  
Best Local Similarity: 49.79% Mismatches: 64  
Query Match: 11.75% Indels: 20  
DB: 7 Gaps: 6

```

US-09-917-376-1 (1-957) x CF876916 (1-747)
Qy 25 SerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAla 44
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 61 TCTCAGATCCTTGGCCCTTGTCTCTGGGGCCGTC-----ATCCCTGCC 102
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Qy 45 HisAlaAlaThrThrGlnProTyThrTrpSerAsnValAlaIle---GlyGlyGlyGly 63
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 103 CATGCTGCC-----TTTTCATGTAAGAACGTCAAGTCTGGCGCGCGCGCGC 150
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Qy 64 PheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyValArgThrAsp 83
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 151 TTCTGCTCCCGCATCATCTTCATCCCAAGACAAAAGGCGTAGCATATGCACGAACAGAT 210
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Qy 84 IleGlyGlyMetTyArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrp 103
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 211 ATTGGCGGCTGTACCGCTCAAC---GCCGACGACTCATGGACCCCGCTCAGGATGGG 267
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Qy 104 Val-----GlyTrpAsnAsnTrpGlyTyAsnGlyValValSerIleAlaAla 119
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 268 ATTGCTGATAATGCGGCTGGCACACTGG-----GGCATCCGACGCTGTTCGCTT 318
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Qy 120 AspProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyThrAsnSerTrpAsp 139
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 319 GATCCGACGACGATCAAAAGGTGTATGCCGAGTCGTCATGTATACGAACAGCTGGAT 378
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Qy 140 ProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrPro 159
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 379 CCGAGTAATGAGCCATCATTCGCTCGTCAGACCGCGCGCAACGTGGTCTTCCACCAAC 438
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Qy 160 LeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaVal 179
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 439 TTGCGCTTCACAGTCGGGGGTAACTGTCAGGACCCGAGCCGAGAGCGTCTGGCTGTC 498
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Qy 180 AspProAsnAsnAspAsnIleLeuTyPheGlyAlaProSerGlyLysGlyLeuTrpArg 199
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 499 GATCCGCCACCTCCACATCATCTACTTTGTGTGCTCGCTCAGACACGCGCTCTGGAAG 558
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Qy 200 SerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTy 219
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 559 TCTACCGAGCGGGCGTGCACCTTTTCCAAAGCGCTCGTCTCAGCGCCCTGGGACGTCC 618
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Qy 220 IleAlaAsnProThrAspThrThrGlyTyrglnSerAspIleGlnGlyValValTrpVal 239
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 619 CTCCCAGACCCGAGTATTCACCGGTACCCAGCGGTACCCAGCGACAGCAAGCCCTCATGTGGTT 678
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Qy 240 AlaPheAspLysSerSerSerLeuGly-GlnAlaSerTyThrIlePheValGlyVa 259
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 679 ACGTTTCGACTCACCCAGCATCAGACCGGTGGGAGCCACCGCTCGTATTTTGTGTCAC 738
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Qy 259 lalaAsp 261
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 739 GACTGAC 745
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::

```

RESULT 4

CB905388

LOCUS

DEFINITION

jecorina cDNA clone tric074xd02, mRNA sequence.

ACCESSION

CB905388

VERSION

CB905388.1 GI:30120046

KEYWORDS

EST.

SOURCE

Hypocrea jecorina (anamorph: Trichoderma reesei)

ORGANISM

Hypocrea jecorina

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.

REFERENCE

AUTHORS

Foreman,P.K., Brown,D.E., Dankmeyer,L., Dean,R., Diener,S.,  
Dunn-Coleman,N.S., Goedegebuur,F., Houfek,T.D., England,G.J.,  
Kelley,A.S., Meerman,H.J., Mitchell,T., Mitchinson,C.,  
Olivares,H.A., Teunissen,P.J., Yao,J. and Ward,M.  
TITLE  
Transcriptional regulation of biomass-degrading enzymes in the

CB905388 814 bp mRNA linear EST 02-JUL-2003  
tric074xd02 T.reesei mycelial culture, Version 3 april Hypocrea  
jecorina cDNA clone tric074xd02, mRNA sequence.

CB905388

CB905388.1 GI:30120046

EST.

Hypocrea jecorina (anamorph: Trichoderma reesei)

Hypocrea jecorina

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.

1 (bases 1 to 814)

Foreman,P.K., Brown,D.E., Dankmeyer,L., Dean,R., Diener,S.,  
Dunn-Coleman,N.S., Goedegebuur,F., Houfek,T.D., England,G.J.,  
Kelley,A.S., Meerman,H.J., Mitchell,T., Mitchinson,C.,  
Olivares,H.A., Teunissen,P.J., Yao,J. and Ward,M.

Transcriptional regulation of biomass-degrading enzymes in the



```

filamentous fungus Trichoderma reesei
J. Biol. Chem. 278 (34), 31988-31997 (2003)
22803314
PUBMED
COMMENT
Contact: Pamela K. Foreman
Genencor Intl.
925 Page Mill Road, Palo Alto, CA 94304, USA
Tel: (650) 846-7635
Fax: (650) 621-7817
Email: pforeman@genencor.com
Seq primer: LT-FI primer.
Location/Qualifiers
1..814
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/stain="QM6a"
/db_xref="taxon:51453"
/cloned_lib="tric074xd02"
/dev_stage="mycelial culture, Version 3 april"
/note="Vector: PREP3Y; Site 1: Not I/Sal I; Mycellal culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

FEATURES
source
Pred. No.: 9.67e-27 Length: 814
Score: 603.50 Matches: 121
Percent Similarity: 65.4% Conservativeness: 38
Best Local Similarity: 49.7% Mismatched: 64
Query Match: 11.75% Indels: 20 Gaps: 6 DB:

ORIGIN
US-09-917-376-1 (1-957) x CB905388 (1-814)

Qy      25 SerPheAlaValAlaAalaLeuGlyValLeuProIIleAlaalleThraLaserProLa 44
|||    :||: |||| |::|   ::|
Dbb     128 TCTCGAGTCTTGCCCTGTGGGGGCCGC-----ATCCTGCC 169
Qy      45 HisAlaAlaThrThrGlnProTyrrTrpSerAsnValAlale---GlyGlyGly 63
|||||  :||: |||| |::|   ::|
Db       170 CATGTTC-----TTTTCATGAAGAACGTTCAACTCGCGCGGGCGGC 217
Qy      64 PheValAspGlylleValPheaSnGlugLyAlaPrOgVileLeutyrrValArgThrAsp 83
|||||  :||: |||| |::|   ::|
Db       218 TTGCCCCGGGATCATCTTCATCCCAGACAAGGGTAGCATATGCGCACAGAT 277
Qy      84 IleGlyMetTyrrArgTrpaspAlaalaaNglyArgrtTpilleProLeuLeuAspTrp 103
|||||  :||: |||| |::|   ::|
Db       278 ATTGGCGGCTGTCGCCTCAAC--GCCGACGACTCATGACC CGCGT CACGATGGG 334
Qy     104 Val-----GlyrTraasnAntRppAlaalavalnglytyrasnglyvalvalserlilaala 119
|||||  :||: |||| |::|   ::|
Db       335 ATGTGTATAATTGCGGCTGGGCACAACTGG-----GGCATCGCGCTGTTCGCTT 385
Qy     120 AspProIleAsnthraasnlvsValrPalaaLavallglyWetTyrrThrAsnSerTrpAsp 139
|||||  :||: |||| |::|   ::|
Db       386 GATCCCAGAGCATCAAAGGTGTATGCCGAGTCGGCATGTATACGNACAGCTGGAT 445
Qy     140 ProAsnAspglylaalleLeuarGsserAspGinglyAlaThrTrpGlnflethrPro 159
|||||  :||: |||| |::|   ::|
Db       446 CCGAGTAATGGACCATCATTCCTCGTCAGACCGCGCGAACCGTGTCTTCCACCAAC 505
Qy     160 LeuProPhelysLeuGlyGlyAsnMetProGlyArarglymetGlyCluaArgLeuAlaVal 179
|||||  :||: |||| |::|   ::|
Db       506 TTGCCCCCTCACATCGCGGGTAACATGTCAGACCGCGAGCGCGAGCGCTCGCTGTC 565
Qy     180 AppProAsnaAsnapnnilleutyrpheGlyAlalaproserylsGlyvleuthraprg 199
|||||  :||: |||| |::|   ::|
Db       566 GATCCCGGCCACCTCCAACAATCATCTTCTTGTGTCTGCTCACGACAGCGGCTCTGGAG 625
Qy     200 SerThrAspservglalyalthrtpsrerginMetThraAsnPheProaspValglyThrtyr 219
|||||  :||: |||| |::|   ::|
```



93	TACGGCACCGGAATGACAATCTTTGGCGGCCACGATCTCACCAACTGGGACACGCGCCAC	152
463	GlnIleHisIleAlaProMetValIysGlyLeuGluThrAlaValAsnAspLeuIle	482
153	AATGTGTCATCAATCACTGGCAGACGGCATCGAGGAATCTTCGTCGAGGACCTGGCC	212
483	SerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHis	502
213	TCTGCACCGCGGGAAGCGAGTATTGGCCGAGTCGGAGACGACACGGCTTCACCTTT	272
503	Ala-----AspValThrAlaValProSerThrIlePheThrSerProValPheThr	519
273	GCCAGCAGAAACGACCTCGGGACATCGCCGACAGCGTCTGGGCAACGCCACATGGGCC	332
520	ThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySer	539
333	ACCTCGACGAGCGTCACTACGCGCGGAACCTCGTCAAGAGCGTGTCCGCGTCGGCAAC	392
540	PheAspProSerSerGlnProAsnArgHisValAlaPheSerThrAspGlyGlyLys	559
393	ACGCGCGGACGCAACAG-----GTGGCCATCTCTGTCGACGCGCGCGCGG	437
560	AsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSer	579
438	ACGTGGAGCATCACTACGCGCGCCGACAGTCCATCAACGCGCGGCGGCTGCTTTCG	497
580	AlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAla	599
498	GCCGACGCGGACACGATCTCTGTGTCGACCGCCCTCGTCCGCGC-----GTGACGCGC	548
600	ValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArg	619
549	TGCGAGTTTTCAGGCGGCGGCTTTGCTCCGTCGAGCGCTGCGCGGCGCGGCTCATCGCC	608
620	SerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSer	639
609	TGCGACAGACCAACAGCGTCTTCTACGCGGCTCCGGATCGACCTTTTACGTCAGC	668
640	ThrAspGlyGlyValThrPhe 646	
669	AAGGACACCGGACGAGCTTC 689	
CB907625 782 bp mRNA linear EST 02-JUL-2003		
tric082xn19 T. reesei mycelial culture, Version 3 april Hypocrea		
jecorina cDNA clone tric082xn19, mRNA sequence.		
CB907625		
CB907625.1 GI:30122283		
EST.		
Hypocrea jecorina (anamorph: Trichoderma reesei)		
Hypocrea jecorina		
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;		
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.		
1 (bases 1 to 782)		
Foreman, P.K., Brown, D.E., Dankmeyer, L., Dean, R., Diener, S.,		
Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., England, G.J.,		
Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchinson, C.,		
Olivares, H.A., Teunissen, P.J., Yao, J., and Ward, M.		
Transcriptional regulation of biomass-degrading enzymes in the		
filamentous fungus Trichoderma reesei		
J. Biol. Chem. 278 (34), 31988-31997 (2003)		
22803314		
12788920		
Contact: Pamela K. Foreman		
Genencor Intl.		
925 Page Mill Road, Palo Alto, CA 94304, USA		
Tel: (650) 846-7635		
Fax: (650) 621-7817		
Email: pforeman@genencor.com		
Seq primer: LT-F1 primer.		
Location/Qualifiers		
1. .782		
FEATURES		
source		
CF882065		
LOCUS		
RESULT 7		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		
CF882065		
LOCUS		



```

DEFINITION   trico29xol3.b11 T.reesei mycelial culture, Version 6 October 2003
ACCESSION    CF882065
VERSION      CF882065.1 GI:38136747
KEYWORDS     EST.
SOURCE       Hypocrea jecorina (anamorph: Trichoderma reesei)
ORGANISM     Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
              Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
REFERENCE    1 (bases 1 to 693)
AUTHORS      Diener, S.E., Dunn-Coleman, N., Foreman, P., Houfek, T.D.,
              Teunissen, P.J.M., van Solingen, P., Dankmeyer, L., Mitchell, T.K.,
              Ward, M. and Dean, R.A.
TITLE        Characterization of the protein processing and secretion pathways
              in a comprehensive set of expressed sequence tags from Trichoderma
              reesei
JOURNAL      FEMS Microbiol. Lett. 230 (2), 275-282 (2004)
COMMENT      Contact: Ralph A. Dean
              Fungal Genomics Laboratory
              North Carolina State University
              Campus Box 7251, Raleigh, NC 27695, USA
              Tel: 919-513-0020
              Fax: 919-513-0024
              Email: ralph.dean@ncsu.edu
              Seq primer: IT-F1 primer.
              Location/Qualifiers
FEATURES     source
              1..693
               /organism="Hypocrea jecorina"
               /mol_type="mRNA"
               /strain="QM6a"
               /db_xref="taxon:51453"
               /clone="trico29xol3"
               /dev_stage="mycelia"
               /clone_lib="T.reesei mycelial culture, Version 6 October
               2003"
               /note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial
               culture grown from 24 hrs to 6 days with varying Carbon
               and Nitrogen sources and concentrations."
ORIGIN
Alignment Scores:
Pred. No.:      1.25e-14      Length:      693
Score:          401.50      Matches:      88
Percent Similarity: 62.30%      Conservative: 26
Best Local Similarity: 48.09%      Mismatches: 49
Query Match:    7.82%      Indels:      20
DB:             7      Gaps:      6

US-09-917-376-1 (1-957) x CF882065 (1-693)

QY 25 SerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAla 44
Db 47 TCTCGAGTCCCTGCTTGTCTCGGCGCGTC-----ATCCCTGCC 88
QY 45 HisAlaAlaThrGlnProTyrThrTrpSerAnValAlaIle---GlyGlyGlyGly 63
Db 89 CATGCTGCC-----TTTTCATGGAGAACGTCAGCTCGGCGCGCGCGCGC 136
QY 64 PheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAsp 83
Db 137 TTGTTCCCGCGCATCATCTTCCATCCCAAGACAAAGCGGTAGCATATGCACGACAGAT 196
QY 84 IleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrp 103
Db 197 ATTGGCGGGTGTACCGCTCAAC---GCCGACGACTCATGCGCGCTCACGATGGG 253
QY 104 Val-----GlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAla 119
Db 254 ATTGCTGATAATGCGCGGTGGCACAACATGG-----GGCATCGCCCTGTTCGGTT 304
QY 120 AspProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAsp 139
Db 305 GATCCGCATGACGATCAAAAGGTGTATGCGCGCATGTATACGACACCTTGAGAT 364

```

```

QY 140 ProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrPro 159
Db 365 CCTACTAATGAGCCCATCATTCCTCGTCAGACACCGCGCAACGTGGTCTCTCACCAAC 424
QY 160 LeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaVal 179
Db 425 TTGCGCTTTCACAGTCAGGCGTAACATGCCAGACTCGGAGCGGAGAGCGTCTGGCTGTC 484
QY 180 AspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeu-TrpAr 199
Db 485 AATCCCTCCAACTACACATCATCTACTGTGGCGTCTGCTCAGAAACAGCTCTCTGGAA 544
QY 199 gSerThr 201
Db 545 GTCTACG 551

RESULT 8
CE342004      748 bp      DNA      linear      GSS 26-SEP-2003
LOCUS        tigr-gss-dog-17000334095340 Dog Library Canis familiaris genomic,
DEFINITION   genomic survey sequence.
ACCESSION    CE342004
VERSION      CE342004.1 GI:36167870
KEYWORDS     GSS.
SOURCE       Canis familiaris (dog)
ORGANISM     Canis familiaris
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE    1 (bases 1 to 748)
AUTHORS      Kirkness, E.F., Bafna, V., Halpern, A.L., Levy, S., Remington, K.,
              Kirsch, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. and
              Venter, J.C.
TITLE        The dog genome: survey sequencing and comparative analysis
JOURNAL      Science 301 (5641), 1898-1903 (2003)
MEDLINE      22875432
PUBMED       14512627
COMMENT      Contact: Kirkness EF
              The Institute for Genomic Research
              Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
              Rockville, MD 20850, USA
              Tel: 301-838-0200
              Fax: 301-838-0208
              Email: ekirknes@tigr.org
              Class: shotgun.
              Location/Qualifiers
FEATURES     source
              1..748
               /organism="Canis familiaris"
               /mol_type="genomic DNA"
               /strain="Standard Poodle"
               /db_xref="taxon:9615"
               /clone_lib="Dog Library"
               /note="Site 1: BstXI; Libraries were prepared from
               peripheral blood"
ORIGIN
Alignment Scores:
Pred. No.:      3.37e-12      Length:      748
Score:          362.00      Matches:      72
Percent Similarity: 83.91%      Conservative: 1
Best Local Similarity: 82.76%      Mismatches: 14
Query Match:    7.05%      Indels:      0
DB:             9      Gaps:      0

US-09-917-376-1 (1-957) x CE342004 (1-748)

QY 782 GlyAlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeuSer 801
Db 194 GGCCACACCTTACCTTACCTTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 253
QY 802 ProSerProSerProSerProSerProSerProSerProSerProSerProSerProSer 821
Db 254 CTTTACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCA 313

```







```

FEATURES
  source      Seq primer: LT-F1 primer.
              Location/Qualifiers
                1..803
                /organism="Hypocrea jecorina"
                /mol_type="mRNA"
                /strain="QM6a"
                /db_xref="taxon:51453"
                /clone="trio16xm14"
                /dev_stage="mycelia"
                /clone_lib="T.reesei mycelial culture, Version 3 april"
                /note="vector: PREP3Y; Site 1: Not I/Sal I; Mycelial
                culture grown from 24 hrs to 6 days with varying Carbon
                and Nitrogen sources and concentrations."
ORIGIN
Alignment Scores:
Pred. No.:      3,96e-12      Length:      803
Score:          361.50      Matches:      88
Percent Similarity: 51.61%      Conservative: 40
Best Local Similarity: 35.48%      Mismatches: 103
Query Match:    7.04%      Indels:      17
DB:             6          Gaps:      6

US-09-917-376-1 (1-957) x CB898982 (1-803)
Qy      453 AsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValIysGly 472
Db      67 CACAGTTTCACCACTCGGACACCGCCACCAATGTGTCATCCATCAATCATCGGACACGCG 126

Qy      473 LeuGluGluThrAlaValAsnAspLeuIleSerProSerGlyAlaProLeuIleSer 492
Db      127 ATCGAGGAATTCTCGTCCAGACCTGGCCCTGACCCTGGCGGAGCGAGCTATTGGCC 186

Qy      493 AlaLeuGlyAspLeuGlyGlyPheThrHisAla-----AspValThrAlaValPro 509
Db      187 GCAGTCGGAGACGACACAGCGCTTCCCTTTGCCAGCAGAAACGACTCGGGACATCGCGC 246

Qy      510 SerThrIlePheThrSerProValPheThrThrGlyThrSerValAspThrAlaGluLeu 529
Db      247 CAGACGGTCTGGGACCCACATATGGGCCACTCGACGAGCGTGCATACCGCGGGAAC 306

Qy      530 AsnProSerIleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArg 549
Db      307 TCGGTCAAGAGGTGCTCGGCTCGGACACCCCGCGGACGACACAG-----354

Qy      550 HisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGly 569
Db      355 ---GTGGCCATCTCGTCCGACGCGCGGCGAGCTGGAGCATCGACTACGCGGCGACACG 411

Qy      570 ValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro 589
Db      412 TCCATGAACGGCGGCGGCGGCGTGTGCGCTATTTCGGCCGACGCGGACACGATCCTCTGTCGAC 471

Qy      590 GlyAspProGlyGlnProValValThrAlaValGlyPheGlyAsnSerTrpAlaAlaSer 609
Db      472 GCCTGTCGCGC-----GTGACGCGCTCGAGTTCCAGGCGAGCTTTTCCTCCGCTC 522

Qy      610 GlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyr 629
Db      523 TCGAGCTCGCGCGGCGGCGGCGCTCATCGCTCGGACAGAGACCAACAGCGCTCTCTTAC 582

Qy      630 AlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProVal 649
Db      583 GCGGCTCGGATCGACCTTTTACGTGACAGAGGACACCGGCGAGCGCTTC-----633

Qy      650 AlaAlaGlyLeuProSerSerGlyAlaValGlyValMet-----PheHisAlaValPro 667
Db      634 ACGCGCGG---CCCAAGCTGGCGAGCGGAGGAGATCGCGGATATCGCTCTCACC 690

Qy      668 GlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGly 687
Db      691 ACCACCGGCGGACGTTGTATGTCGCGGCGGCGGATATTCGCTCCACAGACTCG 750

688 GlySerSerTrpSerAlaIleThr 695
751 GGCAGACCTTTGGCCAAAGTCTCC 774

CNS028QT      791 bp      DNA      linear      GSS 01-SEP-2000
Tetraodon nigroviridis genome survey sequence T7 end of clone
245G14 of library G from Tetraodon nigroviridis, genomic survey
sequence.
AL186302
AL186302.1      GI:7824406
GSS: genome survey sequence.
Tetraodon nigroviridis
Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodontidae; Tetraodon.
1
Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Estimate of human gene number provided by genome-wide analysis
using Tetraodon nigroviridis DNA sequence
Nat. Genet. 25 (2), 235-238 (2000)
20296633
10835645
2
Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
Saurin,W., Bernot,A. and Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Genome Res. 10 (7), 939-949 (2000)
20359837
10899143
3 (bases 1 to 791)
Genoscope.
Direct Submission
Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
  source      Location/Qualifiers
    1..791
    /organism="Tetraodon nigroviridis"
    /mol_type="genomic DNA"
    /db_xref="taxon:99883"
    /clone="245G14"
    /clone_lib="G"
    /note="Genoscope sequence ID : COAG245BD07LP1-end : T7"
ORIGIN
Alignment Scores:
Pred. No.:      4,78e-12      Length:      791
Score:          360.00      Matches:      71
Percent Similarity: 84.88%      Conservative: 2
Best Local Similarity: 82.56%      Mismatches: 13
Query Match:    7.01%      Indels:      0
DB:             9          Gaps:      0

US-09-917-376-1 (1-957) x CNS028QT (1-791)
Qy      783 AlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerPro 802
Db      347 AGCCCTAGCCCTAGCCCTAGCCCTAGCCCTAGCCCTAGCCCTAGCCCTAGCCCT 406

Qy      803 SerProSerProSerProSerProSerProSerProSerProSerProSerProSer 822
```











Db 626 CCGCCAGCGGATGGAGTCATTTCGAACTGTAGCTGCTTCTTTCATTTGTCCCACTGAGG 685  
Qy 894 GlnVal -----ValAsnThrGlySerSerValAspLeu 905  
Db 686 CAAATTTCGATGCAGATTGCACCGGTATGAATGCACACACACACTTGATTTT 745  
Qy 906 SerThrValThrValArgTyrTirPheThrArgAspGlyGlySerSerThrLeuValTyr 925  
Db 746 GGGTCATTGTGTATTTTATTTT-----TCTACAGTTTATAT 784  
Qy 926 AsnCysAsp 928  
Db 785 TG-TGTGAT 792

Search completed: October 5, 2005, 21:09:16  
Job time : 9216.32 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model  
Run on: October 5, 2005, 07:47:30 ; Search time 455.52 Seconds  
(without alignment)  
3437.649 Million cell updates/sec

Title: US-09-917-376-1  
Perfect score: 5135  
Sequence: 1 MDRSENRLTWRRLVSL.....RASFGSVNPATPTADTYLQX 957

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0  
Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO spoel/US09917376/runat 04102005 164328 26925/app query.fasta\_1.2524  
-DB=issued Patents\_NA -QFWT=fastap -SUFFIX=p2n.rn1 -MINMATCH=0 1 -ILOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09917376 @CGN 1.1 516 @runat 04102005 164328 26925 -NCPU=6 -ICPU=3  
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THRAD=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:  
1: /cgn2\_6/ptodata/1/ina/5A COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	409	8.0	11707	3	US-09-136-574A-1
2	358	7.0	19383	4	US-09-949-016-16031
3	345	6.7	18596	3	US-09-318-448-11
4	345	6.7	18597	4	US-09-962-665-8
5	345	6.7	18597	4	US-09-963-333-8
6	345	6.7	18597	4	US-09-962-677-8
7	334	6.5	6416	3	US-09-136-574A-2
8	319.5	6.2	8211	4	US-09-252-991A-13656
9	296.5	5.8	4767	4	US-09-410-551B-28
10	296.5	5.8	4767	4	US-09-940-316B-28
11	293	5.7	4818	4	US-09-410-551B-32
12	293	5.7	4818	4	US-09-940-316B-32

C	13	288.5	5.6	4131	4	US-09-252-991A-13773	Sequence 13773, A
	14	288.5	5.6	4571	4	US-09-410-551B-18	Sequence 18, Appl
	15	288.5	5.6	4571	4	US-09-940-316B-18	Sequence 18, Appl
	16	288	5.6	288	4	US-09-119-507B-111	Sequence 111, App
	17	288	5.6	288	4	US-09-547-693-111	Sequence 111, App
	18	282.5	5.5	441529	3	US-09-103-840A-1	Sequence 1, Appli
	19	281.5	5.5	4403765	3	US-09-103-840A-2	Sequence 2, Appli
	20	280	5.5	3129	4	US-09-252-991A-13873	Sequence 13873, A
	21	279	5.4	4547	4	US-09-410-551B-22	Sequence 22, Appl
	22	279	5.4	4547	4	US-09-940-316B-22	Sequence 22, Appl
	23	273	5.3	4674	4	US-09-410-551B-26	Sequence 26, Appl
	24	273	5.3	4674	4	US-09-940-316B-26	Sequence 26, Appl
	25	269	5.2	4478	4	US-09-410-551B-16	Sequence 16, Appl
	26	269	5.2	4478	4	US-09-940-316B-16	Sequence 16, Appl
	27	268.5	5.2	31467	4	US-09-949-016-13134	Sequence 13134, A
	28	268.5	5.2	31868	4	US-09-949-016-11907	Sequence 11907, A
	29	268	5.2	31391	4	US-09-949-016-14319	Sequence 14319, A
	30	266	5.2	4188	4	US-09-252-991A-13774	Sequence 13774, A
	31	266	5.2	77536	4	US-09-410-551B-1	Sequence 1, Appli
	32	266	5.2	77536	4	US-09-940-316B-1	Sequence 1, Appli
	33	265.5	5.2	77536	4	US-09-410-551B-1	Sequence 1, Appli
	34	265.5	5.2	77536	4	US-09-940-316B-1	Sequence 1, Appli
	35	265	5.2	4737	4	US-09-410-551B-30	Sequence 30, Appl
	36	265	5.2	4737	4	US-09-940-316B-30	Sequence 30, Appl
	37	262	5.1	150394	4	US-09-949-016-13042	Sequence 13042, A
	38	260	5.1	3337	2	US-08-072-610-1	Sequence 1, Appli
	39	260	5.1	3337	2	US-08-719-822B-1	Sequence 1, Appli
	40	260	5.1	3337	3	US-09-092-458-1	Sequence 1, Appli
	41	260	5.1	3337	4	US-08-719-821C-1	Sequence 1, Appli
	42	260	5.1	46253	4	US-09-949-016-11890	Sequence 11890, A
	43	260	5.1	46257	4	US-09-949-016-13711	Sequence 13711, A
	44	258	5.0	4725	4	US-09-410-551B-24	Sequence 24, Appl
	45	258	5.0	4725	4	US-09-940-316B-24	Sequence 24, Appl

ALIGNMENTS

RESULT 1  
US-09-136-574A-1  
; Sequence 1, Application US/09136574A  
; Patent No. 6294366  
; GENERAL INFORMATION:  
; APPLICANT: Farrington, Graham K.  
; Anderson, Paige  
; Gibbs, Moreland  
; Bergquist, Peter  
; Daniels, Roy  
; Morgan, Hugh W.  
; Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for Treating Cellulose Containing Fabrics Using Truncated Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49

CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson

STREET: Spring House Corporate Center, P.O. Box 457

CITY: Spring House

STATE: PA

COUNTRY: USA

ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/136.574A

FILING DATE: 19-Aug-1998

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/932,571

FILING DATE: September 19, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Bak, Mary E.



REGISTRATION NUMBER: 31,215  
 REFERENCE/DOCKET NUMBER: 1997US001/CIP  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 215-540-9200  
 TELEFAX: 215-540-5818  
 TELEX: <Unknown>

## INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 11707 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-136-574A-1

## Alignment Scores:

Pred. No.: 1.52e-11 Length: 11707  
 Score: 409.00 Matches: 263  
 Percent Similarity: 32.37% Conservative: 163  
 Best Local Similarity: 19.98% Mismatches: 386  
 Query Match: 7.96% Indels: 504  
 Gaps: 58

US-09-917-376-1 (1-957) x US-09-136-574A-1 (1-11707)

```

Qy 14 ArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaValAlaAlaLeuGly 33
Db ::::::::::::::::::::-TCCTTATTGTTTTTAAATAAACACCGTGTAGGT 791
Qy 34 ValLeuProIleAlaIleThrAlaSerProAlaHisAlaAlaThrThrGlnProTyrThr 53
Db ::::::::::::::::::::-ATATTCATCAGGAAGCAAGCAGCAGCA-----TATACT 833
Qy 54 TrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsnGluGly 73
Db ::::::::::::::::::::-GTTGATTTTGAAGGT 848
Qy 74 AlaProGlyIleLeuTyr-----ValArgThrAspIleGlyGly 86
Db ::::::::::::::::::::-GCTGATCTTATCTTACTTTGCTTATGGAATAAGTAGTCAGCAGTGTGACATGGGCAAT 908
Qy 87 MetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrp 106
Db ::::::::::::::::::::-GACGTTAAA 983
Qy 107 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 126
Db ::::::::::::::::::::-GCATATAATGGTAAAGTAGTGTGACGGGTGTC-----AATAGAAGTTCA 953
Qy 127 ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 146
Db ::::::::::::::::::::-GACGTTAAA 983
Qy 147 ArgSerAspGlnGlyAlaThrTrpGlnIleThr----- 158
Db ::::::::::::::::::::-AACATATGAAATGAAACCAACATGGGTAGTTTCAGCGTATGTATTAACATAGCTACCAG 1043
Qy 159 ---ProLeuProPheLysLeu-----GlyGlyAsnMetProGlyArg 171
Db ::::::::::::::::::::-AAGCGGTTGCTATTTGGTATCTCAGCGGTTTACGACGATGGAAGTGGGTTAAGAGTACT 1103
Qy 172 GlyMetGlyGluAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 191
Db ::::::::::::::::::::-CTCATAGTGAGGTGTGGCTATT-----CCAAAT----- 1133
Qy 192 ProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 211
Db ::::::::::::::::::::-TATTGGAAGAAAATT--- 1148
Qy 212 AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSer 231
Db ::::::::::::::::::::-GTTGGTAAATGGACTCCAAATATATAGCAATGTGACGAAATTTGTTAATT 1196

```

```

Qy 232 AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGlnAla 251
Db ::::::::::::::::::::- 1241
Qy 252 SerLysThrIlePheValGlyValAlaAspProAsn-----AsnProValPhe 267
Db ::::::::::::::::::::-TATATCCAAATAATGGATGATATAGTTTACCTATCAATGAGTGACA 1289
Qy 268 TrpSerArg-----AspGlyGlyAlaThr-----TrpGlnAla-----Val 279
Db ::::::::::::::::::::- 1349
Qy 280 ProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisVal 299
Db ::::::::::::::::::::- 1394
Qy 300 LeuTyrIleAla-----ThrSerAsnThrGlyGly-----ProTyrAspGly--- 313
Db ::::::::::::::::::::- 1454
Qy 314 ---SerSerGlyAspValTrpLysPheSerVal-----ThrSerGlyThrTrp 328
Db ::::::::::::::::::::- 1514
Qy 329 ThrArgIleSerProValProSerThrAspThrAlaAsnAsp---TyrPheGlyTyrSer 347
Db ::::::::::::::::::::- 1574
Qy 348 GlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrp 367
Db ::::::::::::::::::::- 1634
Qy 368 TrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrp 387
Db ::::::::::::::::::::-TCAATTCCTCGTGGTGTACAGTACGAGTGTGTTG 1670
Qy 388 AspTrpThrSerTyrProAsnArgSerLeuArgTyr-----ValLeu 401
Db ::::::::::::::::::::- 1730
Qy 402 AspIleSer-----AlaGluPro----- 407
Db ::::::::::::::::::::- 1790
Qy 407 ----- 407
Db 1791 GATTATTTCAAAGTAGGAGTAGCTTTGTCTTACAAAAGCATTCGCTCTGTATACAGAAAAG 1850
Qy 407 ----- 407
Db 1851 AAGATGGTTTGAAGCATTTCAATAGTATTACTGTCAGGGAACGAATGAACCATCAGAG 1910
Qy 408 -----TrpLeuThrPheGly 412
Db 1911 TTACTTGTGATGAAATACTTACAACTTTAGCAAGCAGACGAAATTTGTAAATTTTGCA 1970
Qy 413 ValGlnProAsnProProValProSerProLysLeuGlyTrpMetAspGluAla----- 430
Db 1971 ACAAGTAAACAACATTTGCCATCAGAGGTATACACTGGTTTGGCATGAGCAACACCCGAC 2030
Qy 430 ----- 430
Db 2031 TGGTTTTTCAAGGACACAAATGGAATACGTTGAGCAAGGATGCAATGTTAAGCAGATTA 2090
Qy 430 ----- 430
Db 2091 AACAGTATATTTATACGGTAGTGGGAAGATATAAAGGGAAGGTTTATGATCGGATGTG 2150
Qy 431 -----MetAlaIleAspProPheAsnSerAspArgMetLeu----- 442
Db 2151 GTAAATGAAGCAATAGATGAAAGTCAAGGTGATGGATTCAGGAGATCTAACTGTGTACAAC 2210
Qy 443 -----TyrGlyThrGlyAlaThrLeuTyrAla----- 451

```











LOCATION: 15042, 15546, 15770  
OTHER INFORMATION: n = c or t  
NAME/KEY: misc\_feature  
LOCATION: 1322, 1688  
OTHER INFORMATION: n = c or g  
NAME/KEY: misc\_feature  
LOCATION: 2594, 11293, 16199, 16203  
OTHER INFORMATION: n = g or t  
NAME/KEY: misc\_feature  
LOCATION: 3619  
OTHER INFORMATION: n = a or t  
NAME/KEY: misc\_feature  
LOCATION: 14547  
OTHER INFORMATION: nucleotide in position 14547 is t, or absent  
US-09-962-665-8

Alignment Scores:  
Pred. No.: 4, 92e-08 Length: 18597  
Score: 345.00 Matches: 71  
Percent Similarity: 83.91% Conservatives: 2  
Best Local Similarity: 81.61% Mismatches: 13  
Query Match: 6.72% Indels: 1  
DB: 4 Gaps: 0

US-09-917-376-1 (1-957) x US-09-962-665-8 (1-18597)

QY 783 AlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerPro 802  
Db 10388 TCACCATCACCATTCTCCATCACCATCACCATCACCATCACCATCACCATCACCACCA 10329  
QY 803 SerProSerProSerProSerProSerProSerProSerProSerProSerSer 822  
Db 10328 TCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCACCA 10269  
QY 823 SerProSerProSerProSerProSerProSerProSerProSerProSerAla 842  
Db 10268 TCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCACCA 10209  
QY 843 SerProSerProSerProSerProSerProSerProSerProSerProSerSerProSerPr 862  
Db 10208 TCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCACCA 10149  
QY 862 othrProSerSerPro 868  
Db 10148 ATCTCCATCACCATCACCACCA 10130

RESULT 5  
US-09-963-333-8/c  
Sequence 8, Application US/09963333  
Patent No. 6664062  
GENERAL INFORMATION:  
APPLICANT: Stanton, Jr., Vincent P.  
TITLE OF INVENTION: THYMIDINE SYNTHASE GENE SEQUENCE VARIANCES  
TITLE OF INVENTION: HAVING UTILITY IN DETERMINING THE TREATMENT  
TITLE OF INVENTION: OF DISEASE  
FILE REFERENCE: 11926-015002  
CURRENT FILING DATE: 2001-09-24  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 09/596,033  
PRIOR FILING DATE: 2000-06-15  
PRIOR APPLICATION NUMBER: 09/357,743  
PRIOR FILING DATE: 1999-07-20  
PRIOR APPLICATION NUMBER: 09/357,024  
PRIOR FILING DATE: 1999-07-19  
PRIOR APPLICATION NUMBER: 60/093,484  
PRIOR FILING DATE: 1998-07-20  
NUMBER OF SEQ ID NOS: 16  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 8  
LENGTH: 18597  
TYPE: DNA

ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 701, 13751  
OTHER INFORMATION: n = c or a  
NAME/KEY: misc\_feature  
LOCATION: 716, 1293, 2401, 2429, 2618, 3083, 3125, 3635, 4256, 4898,  
LOCATION: 5062, 5167, 11069, 13298, 14479, 14730, 14796, 15344, 15450,  
LOCATION: 15503, 15590, 15840, 16149  
OTHER INFORMATION: n = a or g  
NAME/KEY: misc\_feature  
LOCATION: 732, 1379, 1590, 2488, 3212, 5006, 11238, 11422, 11686,  
LOCATION: 12598, 13171, 13645, 13782, 13806, 13813, 14586, 14788,  
LOCATION: 15042, 15546, 15770  
OTHER INFORMATION: n = c or t  
NAME/KEY: misc\_feature  
LOCATION: 1322, 1688  
OTHER INFORMATION: n = c or g  
NAME/KEY: misc\_feature  
LOCATION: 2594, 11293, 16199, 16203  
OTHER INFORMATION: n = g or t  
NAME/KEY: misc\_feature  
LOCATION: 3619  
OTHER INFORMATION: n = a or t  
NAME/KEY: misc\_feature  
LOCATION: 14547  
OTHER INFORMATION: nucleotide in position 14547 is t, or absent  
US-09-963-333-8

Alignment Scores:  
Pred. No.: 4, 92e-08 Length: 18597  
Score: 345.00 Matches: 71  
Percent Similarity: 83.91% Conservatives: 2  
Best Local Similarity: 81.61% Mismatches: 13  
Query Match: 6.72% Indels: 1  
DB: 4 Gaps: 0

US-09-917-376-1 (1-957) x US-09-963-333-8 (1-18597)

QY 783 AlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerPro 802  
Db 10388 TCACCATCACCATTCTCCATCACCATCACCATCACCATCACCATCACCATCACCACCA 10329  
QY 803 SerProSerProSerProSerProSerProSerProSerProSerProSerSer 822  
Db 10328 TCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCACCA 10269  
QY 823 SerProSerProSerProSerProSerProSerProSerProSerProSerAla 842  
Db 10268 TCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCACCA 10209  
QY 843 SerProSerProSerProSerProSerProSerProSerProSerProSerSerProSerPr 862  
Db 10208 TCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCACCA 10149  
QY 862 othrProSerSerPro 868  
Db 10148 ATCTCCATCACCATCACCACCA 10130

RESULT 6  
US-09-962-677-8/c  
Sequence 8, Application US/09962677  
Patent No. 6759200  
GENERAL INFORMATION:  
APPLICANT: Stanton, Jr., Vincent P.  
TITLE OF INVENTION: THYMIDINE PHOSPHORYLASE GENE SEQUENCE  
TITLE OF INVENTION: VARIANCES HAVING UTILITY IN DETERMINING  
TITLE OF INVENTION: THE TREATMENT OF DISEASE  
FILE REFERENCE: 11926-015003  
CURRENT FILING DATE: 2001-09-24  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 09/596,659



```

; PRIOR APPLICATION NUMBER: 09/596,033
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 09/357,743
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: 09/357,024
; PRIOR FILING DATE: 1999-07-19
; PRIOR APPLICATION NUMBER: 60/093,484
; PRIOR FILING DATE: 1998-07-20
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 18597
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 701..13751
; OTHER INFORMATION: n = c or a
; NAME/KEY: misc_feature
; LOCATION: 716..1293, 2401..2429, 2618, 3083, 3125, 3635, 4256, 4898,
; LOCATION: 5062, 5167, 11069, 13298, 14479, 14730, 14796, 15344, 15450,
; LOCATION: 15503, 15590, 15840, 16149
; OTHER INFORMATION: n = a or g
; NAME/KEY: misc_feature
; LOCATION: 732..1379, 1590, 2488, 3212, 5006, 11238, 11422, 11686,
; LOCATION: 12598, 13171, 13645, 13782, 13806, 13813, 14586, 14788,
; LOCATION: 15042, 15546, 15770
; OTHER INFORMATION: n = c or t
; NAME/KEY: misc_feature
; LOCATION: 1322..1688
; OTHER INFORMATION: n = c or g
; NAME/KEY: misc_feature
; LOCATION: 2594, 11293, 16199, 16203
; OTHER INFORMATION: n = g or t
; NAME/KEY: misc_feature
; LOCATION: 3619
; OTHER INFORMATION: n = a or t
; NAME/KEY: misc_feature
; LOCATION: 14547
; OTHER INFORMATION: nucleotide in position 14547 is t, or absent
US-09-962-677-8

Alignment Scores:
Pred. No.: 4,92e-08 Length: 18597
Score: 345.00 Matches: 71
Percent Similarity: 83.91% Conservative: 2
Best Local Similarity: 81.61% Mismatches: 13
Query Match: 6.72% Indels: 1
DB: 4 Gaps: 0

US-09-917-376-1 (1-957) x US-09-962-677-8 (1-18597)
QY 783 AlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerPro 802
Db 10388 TCACCATCACCATTCCATCTCCATCACCATCACCATCACCATCACCATCACCATCACC 10329
QY 803 SerProSerProSerSerProSerProSerProSerProSerProSerProSerSer 822
Db 10328 TCACCATCTCCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACC 10269
QY 823 SerProSerProSerProSerProSerProSerProSerProSerProSerProSerAla 842
Db 10268 TCACCATCACCATCACCATCTCCATCACCATCACCATCACCATCTCCATCACCATCACC 10209
QY 843 SerProSerProSer-SerSerProSerProSerProSerProSerProSerProSerPr 862
Db 10208 TCACATCACCATTCACCCATCACCATCACCATCACCATCACCATCACCATCTCCATCACC 10149
QY 862 cThrProSerSerPro 868
Db 10148 ATCTCCATCACCATCACCACCA 10130
RESULT 7

US-09-136-574A-2
; Sequence 2, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; ANDERSON, Paige
; GIBBS, Moreland
; BERGQUIST, Peter
; DANIELS, Roy
; MORGAN, Hugh W.
; WILLIAMS, Diane P.
; TITLE OF INVENTION: Compositions and Methods for
; Treating Cellulose Containing
; Cellulase Enzyme Compositions
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Center, P.O. Box 457
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136,574A
; FILING DATE: 19-Aug-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/932,571
; FILING DATE: September 19, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: 1997US001/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6416 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-136-574A-2

Alignment Scores:
Pred. No.: 5,27e-08 Length: 6416
Score: 334.00 Matches: 219
Percent Similarity: 33.27% Conservative: 125
Best Local Similarity: 21.18% Mismatches: 314
Query Match: 6.50% Indels: 376
DB: 3 Gaps: 56

US-09-917-376-1 (1-957) x US-09-136-574A-2 (1-6416)
QY 18 SerLeuLeuAlaAlaThrAlaSerPheAlaValAlaAlaLeuGlyValLeuProIle 37
Db 1189 TCTGTAGTGGCAGAAACA-----GCTGCATCACTTCTGGTGGCTTCAGTT 1233
QY 38 AlalIleThrAlaSerProAlaHisAlaAla----- 47
Db 1234 GTAATTAAGGAAGAATAATCTCAGAAAGCAGCTTCTTATCTCCACATGCCAAAGACCTG 1293
QY 48 -----ThrThrGlnPro-----TyrThrTrpSerAsnValAlaIle 59
Db 1294 TTTGAATTGCGGATACCAACAAGAGTGATGCGGGGTATACTGCTGCACAGGTTTCTAC 1353
QY 60 GlyGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyr 79
```



```
Db 1354 ACATCGGCTGTTTATTGATGACCTTGGATGG-----GCTGCTGATGGCTTAT 1404
Qy 80 ValArgThrAsp-----|||||:||||:||||:|||||
Db 1405 ATTGCGACAAATGACAGTAGTTATTGCGAAAGCTGAAGAGTTGATGTCAGAAATATGCT 1464
Qy 85 GlyGlyMetTyrArgTTP-----|||||:||||:||||:|||||-----Ile 84
Db 1465 AATGGTACTAATACATGACACAATGCTGGGATGATGTTCCGATGGAACATTGATCATG 1524
Qy 100 LeuLeuAspTTPValGlyTTPAsnAsnTTPGlyTyrAsnGlyValValSerIleAlaAla 119
Db 1525 CTTGCAAGATTACAGGGAAGAGTTA-----TATAAGGAGCTGTGGAAAGAACTTA 1578
Qy 120 AspProIleAsnThrAsnLysValTTPAlaAlaValGlyMet-----TyrThrAsnSerTTP 138
Db 1579 GACCATTTGGACTGACAGAAATTACGTATACGCCGAAGGGATGGCATATCTGACAGGATGG 1638
Qy 139 AspProAsnAspGlyAlaIleLeuArgSerSerAspGlnGly-----152
Db 1639 -----GGTTTCATTAGATATCGCAACACAGCTGCATTTTTTAGCATGTGCTAT 1686
Qy 153 AlaThrTTPGln-----|||||:||||:||||:|||||-----IleThr 158
Db 1687 GCAGACTGGTCAGGTGCGATTCGACAAAAGACCAAAATATTTGCACTTTGCCAAAAGC 1746
Qy 159 ProLeuProPheLysLeuGlyGlyAsnMetProGlyArg-----GlyMetGly 174
Db 1747 CAGATTGACTATGCACTGGTTCACCA-----GGTAGAAGTTTGTAGTAGGATTGGC 1800
Qy 175 GluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGly 194
Db 1801 ACCAATTATCCAAACATCCGATCACAGGAAT-----1833
Qy 195 LysGlyLeuTTPArgSerThrAspSerGlyAlaThrTTPSerGlnMetThrAsnPhePro 214
Db 1834 -----CGCATATGTTTCATGGCTAAACAGCATGAAATATACCA 1869
Qy 215 Asp-----ValGlyThrTyrIleAlaAsnProThrAspThrThrGly 228
Db 1870 GAGTATCAGACACATATATATGAGCACTGGTGGTGGTCTCTGGTAGTGATAGT 1929
Qy 229 TyrGlnSerAspIleGlnGlyValValTTP-----ValAlaPheAspLysSerSer 246
Db 1930 TATAATGATGACATTACCGATTATGTACAAATGAGGTTGCTCGCATTAATAATGCTGGA 1989
Qy 247 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal 266
Db 1990 ATTGTTGGTGCACTGGCAAG-----2010
Qy 267 PheTTPSerArgAspGlyGlyAlaThrTTPGlnAlaValProGlyAlaProThrGlyPhe 286
Db 2011 -----ATGTACCAAGTTATATGAGGTGAACCTATTGAT--- 2043
Qy 287 IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThr----- 304
Db 2044 ---GATTTTAAAGCAATTGAACACCCACCAATATGATGAATTTTGTGAATCAAAATTT 2100
Qy 305 SerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTTPLysPheSerValThr 324
Db 2101 GGGNAATTACAGGGTCCA-----AATTATACCGAAGTAATTTTCTATATCTATAAT 2151
Qy 325 SerGlyThrTTPThrArgIleSerProValProSerThrAspThrAlaAsn---AspTyr 343
Db 2152 CGAACAGGATGG-----CCACCAAGGGTAACATGATAAAGTTTAATAT 2199
Qy 344 Phe-----GlyTyrSerGlyLeuThrIleAspArgGln 354
Db 2200 TTATAGACCTAACCGAATTAATCCAGGCGGTAATTCG-----2238
Qy 355 HisProAsnThrIleMetValAlaThrGlnIleSerTTPTrpProAspThrIleIlePhe 374
|||||:||||:||||:|||||
Db 2239 ---CCTGATGTTGTCAAAAGTTGACACA-----TAC 2265
Qy 375 ArgSerThrAspGlyGlyAlaThrTTPThrArgIleTTPAspTTPThrSerTyrProAsn 394
Db 2266 TACATCGAAGGAGGTAAATATTAGCGGTCTCTACGTATGGAC-----AAAAAT 2313
Qy 395 ArgSerLeuArgTyrValLeu---AspIleSerAlaGluProTTPLeuThrPheGlyVal 413
Db 2314 AGGAATATATATACTATGTTCTTGTGGATTTTAGTGAACCAAG-----ATA 2358
Qy 414 GlnProAsnProProValProSerProLysLeuGlyTTP-----MetAspGlu 429
Db 2359 TATCTCTGGCGGTGAAGTTGAACACAAAAGCAGGCTCAATTTAAATATATCTTCCGAC 2418
Qy 430 AlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeu 449
Db 2419 GGGTATCCATGGGATCCTACCAAT---GATCCTCATATAAGGATTAACAGTCATTA 2475
Qy 450 TyrAlaThrAsnAspLeuThrLysTTPAspSerGlyGlyGlnIleHisIleAlaProMet 469
Db 2476 GAAAGAATAAATATATATTCGCCGCATATGATAATAAAT-----CTG 2517
Qy 470 ValLysGlyLeuGluGluThrAlaValAsnAspLeuIleSerPro-----484
Db 2518 GTATGGGTTTAGAGCGCGGTGCGCAACATCCACACCTGCACCAACATCAACACCAACA 2577
Qy 485 -----ProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHis 502
Db 2578 CCAACCCCGCCCAACACCAACAGTGACAGCA-----2610
Qy 503 AlaAspValThrAlaValProSerThrIlePheThr-----SerProValPheThr 519
Db 2611 -----ACGCCGAGCGGACTCTACACCGACACCGAGCGGTCACCT---GGT 2655
Qy 520 ThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleLeValArgAlaGlySer 539
Db 2656 ACGGGAAGTGGTGAAGGTACTGTACAAGCAACATGAGACAAAGTGGCGAGCACAGTTCT 2715
Qy 540 PheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLys 559
Db 2716 ATAAGCGCG-----2724
Qy 560 AsnTTPPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSer 579
Db 2725 ---TGTTTTAAG-----ATAGTGAATGGAGC-----AGCAGCAT 2757
Qy 580 AlaAspGlySerArgPhe-----ValTTP-----AlaProGlyAspProGlyGln 594
Db 2758 GTTGATCTTAGCAGGTTAAGATAAGATACTGTCACACAGTGGATGGTGACAAAGCCACAG 2817
Qy 595 ProValValTyrAlaValGlyPheGlyAsnSerTTPAlaAlaSerGlnGlyValProAla 614
Db 2818 AGTGGCGTATGT-----GACTGG-----2835
Qy 615 AsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGly 634
Db 2836 ---GCACAGATAGGGCAAGCAATGTGACA---TTCATTTTGTGAACCTTAGACGCGGA 2889
Qy 635 ThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuPro 654
Db 2890 GTG-----AGTGGAGCGGATTTATAC---CTGGAGGTAGGATTT---2925
Qy 655 SerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysGlu---GlyAspLeu 673
Db 2926 AGCAGTGGAGCTGGCAGTTTGCAG-----CCTGTAAGGACACAGGGGATATA 2973
Qy 674 TTPLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTTPSerAla 693
Db 2974 CAGGTAAAGGTTTAAACAAGATGACTGGAGCAATTACATTCGGCAGCAGCTGTGTCATGG 3033
Qy 694 IleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSer 713
|||||:||||:||||:|||||
Db 3034 TTCAGAGCATGACGAAT-----TATGAGAGAAATGCGAAGGTGACGCTG 3078
```















Qy	229	---TyrGlnSerAspIleGlnGlyValValTrrpVal-----AlaPheAspLysSer	244
Db	2544	GACTATGCTTCGCACACCCGCGACGCTGAGTACTCCGCGAGAACTACTCGACATCACT	2603
Qy	245	SerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAlaAspProAsnAsn	264
Db	2604	AGGCACAGAGCTCGCAGACCCCGCTCGTGCCTGCTGTGCACCGTGCAGCGGCACC---	2660
Qy	265	ProValPheTrp-----SerArgAspGlyGlyAlaThrTrrpGlnAlaValProGlyAla	282
Db	2661	-----TGGGTCGACAGCCCGCTGGACGGCGAGTACTGGTACCGGAACCTGCGTGAA	2711
Qy	283	ProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeuTyrIle	302
Db	2712	CCGGTCGGTTTCACCCCGCGTCAGCCAGTTCGACGCCCGACCGGCGACACCGTTTCGTC	2771
Qy	303	AlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrrpLysPheSer	322
Db	2772	GAGTCAGCGCCAGCCCGGGTGTTCGAGCGCATGGACGAGATGTCTCAGCGTGGCC	2831
Qy	323	Val-----ThrSerGlyThrTrrpThrArgIleSerProValProSerThrAspThr	339
Db	2832	ACGCTCGTTCGACGACGCGCGACCCCGATG-----CTCACCGCCCTG	2879
Qy	340	AlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIle	359
Db	2880	GCACAGCCCTAT-----GTCCACGGCGTCCACGTGCAC-----	2912
Qy	360	MetValAlaThrGlnIleSerTrrpTrrpAspThrIleIlePheArgSerThrAspGly	379
Db	2913	-----TGCCCGCCATCCTC-----	2927
Qy	380	GlyAlaThrTrrpThrArgIleTrrpAspTrrpThrSerTyrProAsnArgSerLeuArgTyr	399
Db	2928	GGACACACCAACCCGGGTACGGACCTTCGACCTCCGACCTCCACACACCGGTAC	2987
Qy	400	ValLeuAspIleSerAlaGluProTrrpLeuThrPheGlyValGlnProAsnProVal	419
Db	2988	-----TGCTC-----GAGTCGGCTCCCGCGCC	3011
Qy	420	ProSerProLysLeuGlyTrrpMetAspGluAlaMetAlaIleAspProPheAsnSerAsp	439
Db	3012	ACGGCC-----GACTCGGC	3026
Qy	440	ArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrrpAsp	459
Db	3027	CACCCGCTCTCGGCACCGGATCGCTCGCCGG-----	3062
Qy	460	SerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluGluThrAla---Val	478
Db	3063	TGCGCGCGCGGGTTCACGGGTCCCGTGCCTCGCCGCGGACCGCGGTTCATC	3122
Qy	479	AsnAspLeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGly	498
Db	3123	CGCGAAGTGGCGCTCGCGCGCGCGCGCCACC-----GACTCGGC	3164
Qy	499	GlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPhe	518
Db	3165	ACGGTCAACACAGCTCGACGTCCACTCGTGCCTCGCGCGGATCCCGCCCGCAGCGGCC---	3221
Qy	519	ThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGly	538
Db	3222	ACCGCGAGACTGGGTGAT-----	3242
Qy	539	SerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGly	558
Db	3243	-----GAAACCGCGCGCGCGCGCGCGCTTCACCGTCCACACCCGCTCGCGGAC	3296
Qy	559	LysAsnTrrpPheGlnGlySerGluProGlyValThrThrGlyGlyThrValAlaAla	578
Db	3297	GCCCCGTGGACCTGCACCGCGAG-----GGGGTTCTCGCCCGCGCGCGT---CCC	3347

RESULT 10  
US-09-940  
; Sequenc  
; Patent  
; GENERAL



APPLICANT: KOSAN BIOSCIENCES, Inc.  
APPLICANT: REEVES, CHRISTOPHER  
APPLICANT: CHU, DANIEL  
APPLICANT: KHOSLA, CHAITAN  
APPLICANT: SANTI, DANIEL  
APPLICANT: WU, KAI  
TITLE OF INVENTION: POLYKETIDES ENCODING THE FK520 POLYKETIDE SYNTHASE GENE  
FILE REFERENCE: 30062-20026-11  
CURRENT APPLICATION NUMBER: US/09/940,316B  
PRIORITY FILING DATE: 2001-08-27  
PRIORITY FILING DATE: 09/410,551  
PRIORITY FILING DATE: 1999-10-01  
PRIORITY FILING DATE: 1999-06-17  
PRIORITY FILING DATE: 1999-03-11  
PRIORITY FILING DATE: 1998-10-02  
NUMBER OF SEQ ID NOS: 72  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 28  
LENGTH: 4767  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic  
OTHER INFORMATION: PKS synthase fragment  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (3)..(4766)  
US-09-940-316B-28

Alignment Scores:  
Pred. No.: 3,11e-06 Length: 4767  
Score: 296.50 Matches: 233  
Percent Similarity: 35.16% Conservative: 101  
Best Local Similarity: 24.53% Mismatches: 397  
Query Match: 5.77% Indels: 223  
DB: 4 Gaps: 46

US-09-917-376-1 (1-957) x US-09-940-316B-28 (1-4767)

QY 9 LeuThrMetArgSerArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaVal 28  
DB 1827 CTGACCGAACACGAGACCGGCTGCGCCCTACCTGGCGCGGTGCGCGCGGCGGATATA 1886  
QY 29 AlaAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAlaHisAlaAlaThr 48  
DB 1887 CGGCTGTGGCATCGACCGTG-----CGGTGACACCGTGGT-GTTCGAGCACCAGCGC 1939  
QY 49 ThrGlnProTyr----- 52  
DB 1940 CGTACTCTGGAGATGACCGCTCACCGGACCGCGGTGACCGACCGACCGATCGTGT 1999  
QY 53 -----ThrTyrSerAsnValAlaIleGlyGly-----PheVal 65  
DB 2000 TGCTCTTCCCGGCGGCGGTGCGAGTGGCTGGGATGGCGAGTGCACCTGCGCGATTCGTC 2059  
QY 66 AspGlyIleValPheAsnGly-----AlaProGlyIleLeuTyrValArgThrAsp 83  
DB 2060 GGTGTGTTCGCGAGCGATGCGCGAGTGGCGGCGGTGCGCGGAGTTCGT---CGA 2116  
QY 84 IleGlyGlyMetTyrArgTyrAspAlaAlaAsnGlyArgTyrPheLeuLeuAspTyr 103  
DB 2117 CTGGGATCTGTTACGGTTCGATGATCGCGCGGTGG-TGCACCGGTTCATG---TGG 2172  
QY 104 ValGlyTyrAsnAsnTyrGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsn 123  
DB 2173 TCCAGCCC-GCTTCCTGGCGCATG-----ATGGTTTCCTGGCGCG- 2213  
QY 124 ThrAsnLysValTyrAlaAlaValGlyMetTyrThrAsnSer---TyrAspProAsnAsp 142

DB 2214 -----GTGTGGCAGCGCGCGTGTGGCGCGGATGCGGTGATCGGCCATTGCGCAG 2264  
QY 143 GlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGln----- 156  
DB 2265 GGTGAGATCGCGCAGCTTGTGTGGCGGTGCGGTGTCACTACGCGATCCGCCCGCGATC 2324  
QY 157 IleThrProLeuProPheLeuLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArg 176  
DB 2325 GTGACCTTGGCAGCAGCGATCGCGCGGCTGGCGGCGCGGCGCGATGCGCATCC 2384  
QY 177 LeuAlaValAspProAsnAsnAsnIleLeuTyrPheGlyAlaProSerGlyLysGly 196  
DB 2385 :GTCGCCCTG---CCGCGCAGGATGTCAGCTGGTCGACGGGCGC----- 2426  
QY 197 LeuTyrArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspVal 216  
DB 2427 ---TGGATCGCGCGCCACACCGGCGCGCTCCACCGTATCGCGGCGACCCCGGAGCG 2483  
QY 217 GlyThrTyrIleAlaAsnProThrAspThrThrGly----- 228  
DB 2484 GTCGACCATGTCCTCACCGCTCATGAGGCACAAAGGGTGGCGGTGGCGGATCACCGTC 2543  
QY 229 ---TyrGlnSerAspIleGlnGlyValValTrpVal-----AlaPheAspLysSer 244  
DB 2544 GACTATGCTCGCACACCGCGCAGCTCGAGCTGATCCGCGACGAACTACTCGACATCACT 2603  
QY 245 SerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsn 264  
DB 2604 AGCGACAGCAGCTCGCAGACCCCGCTGTCGCGTGGCTGTGACCGTGGACCGCACCC--- 2660  
QY 265 ProValPheTyr-----SerArgAspGlyAlaThrTrpGlnAlaValProGlyAla 282  
DB 2661 -----TGGTTCGACGACCGCGCTGGACGGGAGTACTGTCACGGAACCTGCGTGA 2711  
QY 283 ProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeuTyrIle 302  
DB 2712 CCGGTGCGGTTTCCACCGCGCTGACCGAGTTGACGGCCACGGCGGACACCGGTGTCGTC 2771  
QY 303 AlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSer 322  
DB 2772 GAGGTACGCGCCAGCCCGGTGTTGTGACGGCGATGGACGATGCTCTCAGGTGGCC 2831  
QY 323 Val-----ThrSerGlyThrTrpThrArgIleSerProValProSerThrAspThr 339  
DB 2832 ACGTGTGCTGTGACGACGCGGACCGCCACCGCGATG-----CTCACCGCGCTG 2879  
QY 340 AlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIle 359  
DB 2880 GCACAGGCTAT-----GTCCACGCGGTACCGTCGAC----- 2912  
QY 360 MetValAlaThrGlnIleSerTrpProAspThrIleIlePheArgSerThrAspGly 379  
DB 2913 -----TGGCCCGCATCTCTC----- 2927  
QY 380 GlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyr 399  
DB 2928 GGCACACACACACCGCGGTACTGACCTTCGACCTAGCCCTCAACACCGACCGGTAC 2987  
QY 400 ValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProVal 419  
DB 2988 -----TGGCTC-----GAGTCGCGTCCCGCGGC 3011  
QY 420 ProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAsp 439  
DB 3012 ACGGCC-----GACTCGGCG 3026  
QY 440 ArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAsp 459  
DB 3027 CACCCCGCTCGCGACCGAGTCCGCTCGCGCG- 3062  
QY 460 SerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAla-----Val 478  
DB 3063 TCGCGGCGCGGTGTTACGCGGTCCCGTGGCGCGGTGCGGACCGCGGTGTTTCATC 3122



QY 811 erProSerPro-----SerProSerSerSerProSerSerSerProS 825  
DB 4076 CCACCAACCCACCTACGCTCACCACCAACACCTCCACACCCACCTACCCCTC 4135  
QY 825 erProSerProSerProSerProSerProSerProSerProSerProS 845  
DB 4136 CACCAACCCACCAACACCAACCAACCCACCCACCCCTCAACCCCAACCA 4195  
QY 845 erProSerSer-----ProSerProSerSerProSerSerSerProS 860  
DB 4196 CCGCATCTCTATCACCAGGGGCTCGGCACCTCTCGCGCATCTCTCGCGCACCTCAA 4255  
QY 860 roSerProThrProSerSerPro 868  
DB 4256 CCACCCACCACTACCTCTCTCTCC 4281  
RESULT 11  
US-09-410-551B-32  
; Sequence 32, Application US/09410551B  
; Patent No. 6503737  
; GENERAL INFORMATION:  
; APPLICANT: KOSAN BIOSCIENCES, Inc.  
; APPLICANT: REEVES, CHRISTOPHER  
; APPLICANT: CHU, DANIEL  
; APPLICANT: KHOSLA, CHAITAN  
; APPLICANT: SANTI, DANIEL  
; APPLICANT: WU, KAI  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA  
; FILE REFERENCE: 30062-20026.00  
; CURRENT APPLICATION NUMBER: US/09/410,551B  
; PRIOR FILING DATE: 1999-10-01  
; PRIOR APPLICATION NUMBER: US 60/139,650  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/123,810  
; PRIOR FILING DATE: 1999-03-11  
; PRIOR APPLICATION NUMBER: US 60/102,748  
; PRIOR FILING DATE: 1998-10-02  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 32  
; LENGTH: 4818  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURES:  
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic  
; OTHER INFORMATION: PKS synthase fragment  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (3)..(4817)  
US-09-410-551B-32  
Alignment Scores:  
Pred. No.: 4,76e-06 Length: 4818  
Score: 293.00 Matches: 235  
Percent Similarity: 33.07% Conservative: 97  
Best Local Similarity: 23.41% Mismatches: 379  
Query Match: 5.71% Indels: 295  
DB: 4 Gaps: 48  
US-09-917-376-1 (1-957) x US-09-410-551B-32 (1-4818)  
QY 1 MetAspArgSerGluAsnIleArgLeuThrMetArgSer----- 13  
DB 1798 CTGACGCGCTCCCGCGCGCGCGCTCAGCAGCCGCGCGAAGACCTTCGCTGCTCGTGT 1857  
QY 14 ArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaValAlaAlaLeuGly 33  
DB 1858 CGGCGGCTTCCCGGAGGCACCTCGACGAGATCGGCGCTCGCGC----- 1905  
QY 34 ValLeuProfileAlaIleThrAlaSerProAlaHisAlaAlaThrThrGlnProThr 53  
DB 1906 -----CCTATC-----TCGACACCGCGCGCGCGCTCGACCGCGCGCTGGCGC 1950

QY 479 AsnAspLeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGly 498  
DB 3123 CCGCACTGGCGCTCGCGCGCGCGACCGCACC-----GACTGGCGCC 3164  
QY 499 GlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPhe 518  
DB 3165 ACGGTGACAGCTCGAGTCACTCTCGTGGCGCGGATCGCGCGCGGAGGCGC--- 3221  
QY 519 ThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleValValArgAlaGly 538  
DB 3222 ACCGCGCAGACTGGGTGAT----- 3242  
QY 539 SerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGly 558  
DB 3243 -----GAACCGCGCGCGCGCGCGCGCTTCCACCGTCCACACCGCGGTGCGCGAC 3296  
QY 559 LysAsnThrPheGlnGlySerGluProGlyValThrThrGlyGlyThrValAlaAla 578  
DB 3297 GCCCGTGGACGCTGCACCGCGAG-----GGGGTCTCCGCGCGCGCGGTG---CCC 3347  
QY 579 SerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyr 598  
DB 3348 CAGCCCGAAGCGCTGCACCGCGCTGCGCCCG-----CGGCGC-----CGCGTGGCC 3395  
QY 599 AlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIle 618  
DB 3396 GCGGACGGCTGCGCGCGCGTGGCGCGCGGACCGAGTCTTCTGTCGAAGCCGAAGTC 3455  
QY 619 ArgSerAspArgValAsnProIleThrPheTyrAla-----LeuSerAsnGlyThr 635  
DB 3456 GACAGC-----CCTGACCGCTTCTGGGACACCCCGACCTGCTCGACCGCGGTG 3503  
QY 636 PheTyrArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSer 655  
DB 3504 TTCTCCGGTGGCGGACGG-----AGCCGCGAGCGGACCGGATG-CGCGACCTCGC 3556  
QY 656 SerGlyAlaValGlyValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeu 675  
DB 3557 GGTGCACGCTGGGACCGCACCGTGTGCGCGCTGCTCACCCTCGCGCGCGACAGTGTGT 3616  
QY 676 AlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThr 695  
DB 3617 CGTGGAGCTCGCGCTTTCGACCGTCCGGAATGCGGT-----GCTCACCGC 3664  
QY 696 GlyValSerSerAla-----ValAsnValGlyPheGlyLysSerAlaProGlySerSer 713  
DB 3665 GGAGTGGTGGCGGAGTGGCGGAGTGGCGGAGGCGGATCGGACGAGTGGACGGTCT 3724  
QY 714 TyrProAlaValPheValValGlyThrIleGlyGlyVal-ThrGlyAlaTyrArgSerAs 733  
DB 3725 GCTTCGGCTTGAGTGTGCTGCGGTGGCGAGGCCACTACGACGGTGCC-----GA 3775  
QY 733 pAspCysGlyThrThrTrpValLeuIleAsn-----AspAspGlnHisGlnTyr 749  
DB 3776 CGAGTGGCGGAGGCTACACCTCTATCACCAGCCACACACCCCGACGACCCCGCGACCC 3835  
QY 749 rGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArgVal----- 766  
DB 3836 CACCAACCCCGCAACACACACCGACCGACCCACCAACACACACCGGTCTCTACCGC 3895  
QY 767 -----TyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle----- 780  
DB 3896 CCTCCACACCACTCATCACCACCAACACACCTCTATCGTCCACACCGACCGACCC 3955  
QY 781 ----GlyGlyAlaProSerGly-SerProSerPro-----SerValSerProSerAlaAs 797  
DB 3956 CCGAGGCGCGCGCTCACCAGGCTCACCAGCGACCGCACAAACGAAACACCCCGCGCGCAT 4015  
QY 797 erProSerLeuSerPro-----SerProSerProSerSerProS 811  
DB 4016 CCACCTCATCGAAACCCACACCCCGCACCGCGCGCGCTCACCCTCACCCTACCACTC 4075



QY 54 -----TrpSerAenValAlaIleGlyGlyGlyPheValAspGlyIleValPheAen 71  
 Db 1951 AGACACTGG----- 1959  
 QY 72 GluGlyAlaProGlyIleLeuTyValArgThrAspIleGlyMetTy- 88  
 Db 1960 -----CCCGGC-----GTACGCACCTTACCACCGGCGGTACTGTCTCGGG 2001  
 QY 88 ----- 88  
 Db 2002 ACACCGTCATCGCGCTCCCCCGGAGCAGCGCGAGCAACTGCTTCTCTACTCCG 2061  
 QY 89 -----ArgTTPAspAla 92  
 Db 2062 GTCAGGGCACCAGATCCCGCGATGGCGGAGCGAGCTAGTGGTGGTGGTTCG 2121  
 QY 93 AlaAenGly----- 95  
 Db 2122 CCGAGCGGATGGCCGAGTGTGGCGGGTGGCGGAGTTCGTGGACTGGGATCTGTTC 2181  
 QY 96 -----ArgTTPleProLeuLeuAspTTPValGlyTTPAsnAsnTTP 109  
 Db 2182 CGGTTCTGGATGATCCCGCGGTGGTGGACCGGGTTGATGTGTCAGGCC-CTTCTCTGG 2240  
 QY 110 GlyTyAenGlyValValSerIleAlaAlaAspProIleAenThrAsnLysValTTPAla 129  
 Db 2241 GCGATG-----ATGTTTCTCTGCGCG-----GTGTGGCAG 2273  
 QY 130 AlaValGlyMetTyThrAsnSer---TTPAspProAsnAspGlyAlaIleLeuArgSer 148  
 Db 2274 GCGCGGTGTGGCGCGGATGGTGTGATCGGCATTCGCGGATGAGATCGCCGAGCT 2333  
 QY 149 SerAspGlnGlyAlaThrTTPGln-----IleThrProLeuProPhe 162  
 Db 2334 TGTGTGGCGGTGGTGTCTACTAGCGATGCGCGCGGATCGTGCACCTTGGCGCCAG 2393  
 QY 163 LysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAen 182  
 Db 2394 CGCATCGCCCGCGGCGCGCGCGCGCGGATGGATGCGATCGCCCTG---CCCGCG 2450  
 QY 183 AsnAspAenIleLeuTyPheGlyAlaProSerGlyLysGlyLeuTTPArgSerThrAsp 202  
 Db 2451 CAGGATGTGAGTGGTGTGCAGCGGCGC-----TGGATCGCCCGCCAC 2492  
 QY 203 SerGlyAlaThrTTPSerGlnMetThrAenPheProAspValGlyThrTyIleAlaAen 222  
 Db 2493 AAGCGCGCGCTCCACCGTGTATCGCGGCGCGCGCGGATGGATGCGATCTCTCACC 2552  
 QY 223 ProThrAspThrGly-----TyrGlnSerAspIle 233  
 Db 2553 GCTCATGAGGCACAAAGGGGTGGGGTGGCGGATCACCGTTCGACTATGCTCTCGCACACC 2612  
 QY 234 GlnGlyValValTTPVal-----AlaPheAspLysSerSerSerLeuGlyGln 250  
 Db 2613 CGGACGTGAGTGTATCGGACGAACTACTCGATCATCTAGCAGCAGAGCTCGCAG 2672  
 QY 251 AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTTP----- 268  
 Db 2673 ACCCGCTCGTGGTGGTGTGCGACCGTGGAGCGGCGC-----TGGGTTCGAC 2720  
 QY 269 SerArgAspGlyGlyAlaThrTTPGlnAlaValProGlyAlaProThrGlyPheIlePro 288  
 Db 2721 AGCCCGCTGGAGGGGAGTACTGTGTACCGGAACCTCGGTGAACCGGTTCGTTTCCACCCC 2780  
 QY 289 HisLysGlyValPheAspProValAenHisValLeuTyIleAlaThrSerAsnThrGly 308  
 Db 2781 GCCGTGACCGATTGACGCCCGGCGGACACCGGTGTCGTGAGGTGAGCGCCAGCCCG 2840  
 QY 309 GlyProTyAspGlySerSerGlyAspValTTPLysPheSerVal-----ThrSer 325  
 Db 2841 GTGTTGTTCAGCGCATGGACACGATGTCTGTACCGGTGGCCACCGTGTGTCGTGACGAC 2900

QY 326 GlyThrTTPThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGly 345  
 Db 2901 GCGCAGCGCCACCGGATG-----CTCACCGCGCTGGCACAGGCTAT----- 2942  
 QY 346 TyrSerGlyLeuThrIleAspArgGlnHisProAenThrIleMetValAlaThrGlnIle 365  
 Db 2943 GTCCACGCGGTACCGGTGCAC----- 2963  
 QY 366 SerTTPTTPAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTTPThrArg 385  
 Db 2964 -----TGGCCCGCATCTCTC-----GGCACCCACACACCGCG 2996  
 QY 386 IleTTPAspTTPThrSerTyProAsnArgSerLeuArgTyValLeuAspIleSerAla 405  
 Db 2997 GTACTGGACCTTCGACCTTCCACACACCGCGGTAC----- 3038  
 QY 406 GluProTTPLeuThrPheGlyValGlnProAenProProValProSerProLysLeuGly 425  
 Db 3039 -----TGGCTC-----GAGTCGGCTCCCGCGCCACCGGCC----- 3068  
 QY 426 TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyGlyThr 445  
 Db 3069 -----GACTCGGGCCACCCCGTCTCGGCAC 3095  
 QY 446 GlyAlaThrLeuTyAlaThrAsnAspLeuThrLysTTPAspSerGlyGlyGlnIleHis 465  
 Db 3096 GGAGTCGCGGTGCGCGG-----TCGCGCGCGCGGTGTT 3131  
 QY 466 IleAlaProMetValLysGlyLeuGluThrAla---ValAsnAspLeuIleSerPro 484  
 Db 3132 ACGGTGCGGTGCGCGCGGTGCGGCGGTGTTTCATCGCCGAACTGGCGGTGCGC 3191  
 QY 485 ProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAsp 504  
 Db 3192 GCGCGCGCGCCACCC-----GACTCGCCACCGTTCGACAGCTCGAC 3233  
 QY 505 ValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerVal 524  
 Db 3234 GTCACTCGGTGCGCGCGGTGCGCGCGCGCGGCGC---ACCGCGACAGCTGGGTG 3290  
 QY 525 AspyTyAlaGluLeuAsnProSerIleValArgAlaGlySerPheAspProSerSer 544  
 Db 3291 GAT-----GAACCGCGCGC 3305  
 QY 545 GlnProAsnAspArgHisValAlaPheSerThrAspGlyLysAsnTTPPheGlnGly 564  
 Db 3306 GACGCGCGCGCGGTTCACCGTCCACCGCGTGGCGCGCGCGCGCGCGCGCGCGC 3365  
 QY 565 SerGluProGlyGlyValThrThrGlyThrValAlaAlaSerAlaAspGlySerArg 584  
 Db 3366 GCGGAG-----GGGTTCTCGCGCGCGCGCGGTG---CCCCAGCCGAGCGGTG 3416  
 QY 585 PheValTTPAlaProGlyAspProGlyGlnProValTyAlaValGlyPheGlyAsn 604  
 Db 3417 ACCGCTGCGCGCGC-----CCGCGC-----GCGGTGCGCGCGCGCGGTG 3464  
 QY 605 SerTTPAlaAspSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsn 624  
 Db 3465 GCGTGGCGACGCGCGCGCGGTCTTCGTGGAAGCCGAGTGCACAGC----- 3512  
 QY 625 ProLysThrPheTyAla-----LeuSerAsnGlyThrPheTyArgSerThrAsp 641  
 Db 3513 CTGACGGCTTCTGTGGCACACCCGACCTGTCGACGCGGTCTTCTCCGCGGTGCGCGC 3572  
 QY 642 GlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyVal 661  
 Db 3573 GGG-----AGCGCGCAGCGCGCGGATG-GCGCGACCTCGCGGTGCGCGGTGCGC 3625  
 QY 662 MetPheHisAlaValProGlyLysGlyGlyAspLeuTTPLeuAlaAlaSerSerGlyLeu 681  
 Db 3626 CACCGTGTGCGCGCGCTGCTCACCGCGCGCGACAGTGTGTGTGTCGTGAGCTCGCGC 3685  
 QY 682 TyrHisSerThrAsnGlySerSerTTPSerAlaIleThrGlyValSerSerAla--- 700



```

; SEQ ID NO 32
; LENGTH: 4818
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic
; OTHER INFORMATION: PKs synthase fragment
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)..(4817)
US-09-940-316B-32

Alignment Scores:
Pred. No.: 4,76e-06 Length: 4818
Score: 293.00 Matches: 235
Percent Similarity: 33.07% Conservative: 97
Best local Similarity: 23.41% Mismatches: 379
Query Match: 5.71% Indels: 295
DB: 4 Gaps: 48

US-09-917-376-1 (1-957) x US-09-940-316B-32 (1-4818)
QY 1 MetAspArgSerGluAsnIleArgLeuThrMetArgSer----- 13
Db 1798 CTGGACCGCTCCCGCGCGCGCGCTCAGACACCGCGGCGAAGACCTTCCGCTGCTCGTGT 1857
QY 14 ArgArgLeuValSerLeuAlaAlaThrAlaSerPheAlaValAlaAlaLeuGly 33
Db 1858 CGCGCGGTTCCTCCCGGAGGCACTCGACGACGATCGGCGCGCTGCGCG----- 1905
QY 34 ValLeuProIleAlaIleThrAlaSerProAlaHisAlaAlaThrThrGlnProTyrThr 53
Db 1906 -----CCTATC-----TCGACACCGCGCGCGCTCGACCGCGCGCGCTGCGCGC 1950
QY 54 -----TrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 71
Db 1951 AGACACTGG----- 1959
QY 72 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyr----- 88
Db 1960 -----CCGGC-----GTACGCACTTCCACCCGCGCGGTACTGCTCGGGG 2001
QY 88 ----- 88
Db 2002 ACACCGTCATCGCGCTCCCGCGGACGAGCCGACGAACTCGTCTTCTACTCTCCG 2061
QY 89 -----ArgTrpAspAla 92
Db 2062 GTACGGGCACCCAGCATCCCGCGATGGCGGAGCAGCTAGCCGATTCTCGTGGTGTTCG 2121
QY 93 AlaAsnGly----- 95
Db 2122 CCGAGCGGATGGCGGAGTGTGCGCGCGGTTCGCGAGTTCGTGGACTGGGATCTGTTC 2181
QY 96 -----ArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrp 109
Db 2182 CGGTCTCGATGATCCGCGGCTGTGGACCGGGTGTATGTGGTCCAGCCC-GCTTCTCG 2240
QY 110 GlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnIleValTrpAla 129
Db 2241 GCGATG-----ATGGTTTCCCTGCGCGG-----GTGTGGCAG 2273
QY 130 AlaValGlyMetTyrThrAsnSer---TrpAspProAsnAspGlyAlaIleLeuArgSer 148
Db 2274 GCGGCGGTGTGGCGCGGATGCGGTGATCGGCCATTTCGAGGGGTGAGATCGCGCAGCT 2333
QY 149 SerAspGlnGlyAlaThrTrpGln-----IleThrProLeuProPhe 162
Db 2334 TGTGTGCGGCTGCGGTGTCTACACGCGATGCGCGCGCGGATCGGTGACCTTGGCGGACCG 2393
QY 163 LysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsn 182
Db 2394 GCGATCGCGCGGCTGCGGCGCGCGCGCGATGCGATCGCTCGCGCTG---CCGCG 2450
```

```

Db 3686 CGAGGTGCCGATGCCGT-----GCTACCCGCGAGTCTGCTGCGCTGG 3733
QY 701 ---ValAsnValGlyPheGlyLysSerAlaProGlySerTyrProAlaValPheVal 719
Db 3734 CGAGGTGCGTCCGCGGCGGATCGAGAGTCCGACGCTCTGCTTCGCTTGTAGTGTGT 3793
QY 720 ValGlyThrIleGlyGlyVal-ThrGlyAlaTyrArgSerAspAspCysGlyThrThr 739
Db 3794 GCGGTGCGGAGGCCACTACGAGGTGCC-----GACGAGCTGCCGCGAGGGCTA 3844
QY 739 pValLeuIleAsn-----AspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAl 755
Db 3845 CACCTCTATACCGCCACACACACCCCGACGCCCGACGCCACCCCAACCCCAACAC 3904
QY 755 aileThrGlyAspHisAlaAsnLeuArgVal-----TyrIle 768
Db 3905 ACCACACGACCCACACACACAAACACACGCGTCTCAGCGCCCTCCCAACACCACTCAT 3964
QY 768 eGlyThrAsnGlyArgGlyIleValTyrGlyAspIle-----GlyGlyAlaProSe 785
Db 3965 CACCAACCAACACACCTCATGCTCCACACACCAACCGACCCCGCGCGCGCTCAC 4024
QY 785 rGly-SerProSerPro-----SerValSerProSerAlaSerProSerLeuSerPro- 802
Db 4025 CGGCTCACCAGCCACCGCACAAACGAAACACACCGCGCGCATCTCACCTCATCGAAACCA 4084
QY 803 -----SerProSerProSerSerProSerProSerPro----- 814
Db 4085 CCACCCCNACACCCACTCCCTCACCACCTCACCACCTCCACCAACCCCACTAGC 4144
QY 815 -----SerProSerSerProSerSerProSerProSerProSerProSerProS 831
Db 4145 CCTCACCAACACACCTCCACACCCCGCACCTCACCCTCATCACCACCCCAACAC 4204
QY 831 erProSerProSerArgSerProSerProSerProSerProSerProSerProSerS 849
Db 4205 CACCAACACCCCAACACCCCGCACCTCACCACCCCAACCAACCGCGCTCATCACC 4264
QY 850 -----ProSerProSerSerProSerProSerProSerProSerProSerProS 866
Db 4265 CGCTCGCGCACCTCGCGCGCATCTCTCGCGCGCATCTCACCACCAACCCCACTACT 4324
QY 866 erSerPro 868
Db 4325 CCTCTCCC 4332

RESULT 12
US-09-940-316B-32
; Sequence 32, Application US/09940316B
; Patent No. 6759536
; GENERAL INFORMATION:
; APPLICANT: KOSAN BIOSCIENCES, Inc.
; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTIL, DANIEL
; APPLICANT: WU, KAI
; TITLE OF INVENTION: POLYPEPTIDES ENCODING THE FKBP GENE OF THE FK-520 POLYKETIDE SYNTH
; FILE REFERENCE: 30062-20026.11
; CURRENT FILING DATE: 2001-08-27
; PRIOR FILING DATE: 1999-10-01
; PRIOR FILING DATE: 1999-06-17
; PRIOR FILING DATE: 1999-03-11
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
```



```
QY 183 AsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAsp 202
Db 2451 CAGGATTCGAGCTGGTCGACGGGGCC-----TGGATCGCGCCGAC 2492
QY 203 SerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsn 222
Db 2493 AACGGGCGCGCTCCACCGTGATCGCGGACCCCGGAGCGGTGCACCATCTCTCACC 2552
QY 223 ProThrAspThrThrGly-----TyrGlnSerAspIle 233
Db 2553 GCTCATGAGGCACAAGGGGTGGGTGGCGGATCACCGTCGACTATGCTCGCACACC 2612
QY 234 GlnGlyValValTrpVal-----AlaPheAspLysSerSerSerLeuGlyGln 250
Db 2613 CCGCAGCTCGAGCTGATCCGCGAGAACTACTCGACATCACTAGCGACAGAGTCTCGCAG 2672
QY 251 AlaSerLysThrIlePheValGlyValAlaAspProAsnProValPheTrp----- 268
Db 2673 ACCCGCTGCTGGCTGGCTGACCGTGCACGGCACC-----TGGGTGCAC 2720
QY 269 SerArgAspGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIlePro 288
Db 2721 AGCGCGTGGACGGAGTACTGGTACCGGAACCTCGGTGAACCGGTTCGTTCCACGCC 2780
QY 289 HisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGly 308
Db 2781 GCCGTACGCCAGTTGACAGCCCGAGCGGACACCGCTGTCGAGGTGACGCCGCCGCG 2840
QY 309 GlyProTyrAspGlySerSerGlyAspValTrpLysPheSerVal-----ThrSer 325
Db 2841 GTGTTGTGACGGGATGACGACGATGCTCAGGTGTCACGCTGCGTGTGTCGACGAC 2900
QY 326 GlyThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGly 345
Db 2901 GCGCAGCGCACCGCGATG-----CTCACCGCGCTGGCACAGCGCTAT----- 2942
QY 346 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValaThrGlnIle 365
Db 2943 GTCCACGGCGTACCGTCTGAC----- 2963
QY 366 SerTrpTrpProAspThrIlePheArgSerThrAspGlyGlyAlaThrTrpThrArg 385
Db 2964 -----TGCGCGCGCATCCTC-----GGCACCCACACACCGCG 2996
QY 386 IleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAla 405
Db 2997 GTACTGGACCTCCGACCTACCGCTTCCACACACCGCGGTAC----- 3038
QY 406 GluProTrpLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGly 425
Db 3039 -----TGCGTCT-----GAGTCGGCTCCCGCGCACCGGCC----- 3068
QY 426 TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr 445
Db 3069 -----GACTCGGGCCACCCCGTCTCTCGGCACC 3095
QY 446 GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHis 465
Db 3096 GGAGTCGGCTCGCGCG-----TCGCGGGCGGGGTGTC 3131
QY 466 IleAlaProMetValLysGlyLeuGluThrAla-----ValAsnAspLeuIleSerPro 484
Db 3132 ACGGGTCCGCTGCGCGCGGTGCGGACCGCGCGGTGTTTCATCGCGCGGTGTCGCGCTGCC 3191
QY 485 ProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAsp 504
Db 3192 CGCGCGGACGCCACC-----GACTCGCGCACCGGTGCGAAGCAGTCTGCAC 3233
QY 505 ValThrAlaValProSerThrIlePheThrSerProValPheThrGlyThrSerVal 524
Db 3234 GTACCTCGGTGCGCGGATTCGCCCGCGCGGAGGGCC-----ACCGCGCAGACCTGGGTCTC 3290
```

```
QY 525 AspTyrAlaGluLeuAsnProSerIleValArgAlaGlySerPheAspProSerSer 544
Db 3291 GAT-----GAACCGCGCGCC 3305
QY 545 GlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGly 564
Db 3306 GACGGCGCGCGCTTCCACCGTCCACACCGCGTCCGCGACCGCGCGTGGACGCTGCAC 3365
QY 565 SerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArg 584
Db 3366 CCGGAG-----GGGTTCCTCGCGCGCGCGCGGTG-----CCCCAGCCGACCGTGCAG 3416
QY 585 PheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsn 604
Db 3417 ACCGCTGGCGCGCG-----CGGGC-----GCGGTGCCCGCGGCGGTGCCCGGG 3464
QY 605 SerTrpAlaAspSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsn 624
Db 3465 GCGTGGCGACGCGCGGACGAGTCTTCGTCGAAGCGGAAGTCGACAGC----- 3512
QY 625 ProLysThrPheTyrAla-----LeuSerAsnGlyThrPheTyrArgSerThrAsp 641
Db 3513 CTTGACGGCTTCTGTCGACACCGCGACCTGTCGACGGGTCTTCTCCGCGTTCGCGCAG 3572
QY 642 GlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyVal 661
Db 3573 GGG-----AGCGCGCAGCGCACCGGATG-GCGCGACCTCGCGGTGCGACCGTGCAGCGC 3625
QY 662 MetPheHisAlaValProGlyLysGlyAspLeuTrpLeuAlaAspSerSerGlyLeu 681
Db 3626 CACCGTGTGCGCGCTCCTCACCGCGCGACAGTGTGTCGTGAGAGTCTCGCGCTT 3685
QY 682 TyrHisSerThrAsnGlySerSerTrpSerAlaIleThrGlyValSerSerAla--- 700
Db 3686 CGACGGTCCCGAATGCGGT-----GCTCACCGCGGAGTCTCGTTCGCGTTCGAGTGGT 3733
QY 701 ---ValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheVal 719
Db 3734 CGAGTCCGTCGCGGCGGATCCGACGATCGGACGCTGCTGCTTCGCTTCGAGTGGTT 3793
QY 720 ValGlyThrIleGlyGlyVal-ThrGlyAlaTyrArgSerAspCysGlyThrThrTr 739
Db 3794 GCGCGTGGCGGAGGCCACTACGCGGTGCC-----GACGAGTCTCCCGGAGGCTA 3844
QY 739 pValleuIleAsn-----AspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAl 755
Db 3845 CACCTCATACCGCGCACACACCGCGACCGACCGACCGACCGACCGACCGACCGAC 3904
QY 755 aIleThrGlyAspHisAlaAsnLeuArgVal-----TyrIle 768
Db 3905 ACCCACGCGACCCACACACACACACACGCGTCTCTACCGCGCTCCACACACCGCTCAT 3964
QY 768 eGlyThrAsnGlyArgGlyIleValTyrGlyAspIle-----GlyGlyAlaProse 785
Db 3965 CACACACACACACACCTCATCTGTCACACACCGACCGACCGACCGCGCGCGCTGCAC 4024
QY 785 rGly-SerProSerPro-----SerValSerProSerSerAlaSerProSerLeuSerPro- 802
Db 4025 GCGGCTCACCGCGACCGCACAAACGAAACACCGCGCGCGCTCATCTCATCGAAACCCA 4084
QY 803 -----SerProSerProSerSerSerProSerProSerProPro----- 814
Db 4085 CCACCCCGACACCCCGCTCCCGCTCACCACTCACCACTCCCGCTCCCGCTCCCGCTCCG 4144
QY 815 -----SerProSerSerProSerSerProSerProSerProSerProSerProS 831
Db 4145 CCTCACCAACACACCTCCACACCGCGCTCATCTCCCGCTCCCGCTCCCGCTCCCGCT 4204
QY 831 erProSerProSerArgSerProSerProSerAlaSerProSerProSerSerSer----- 849
Db 4205 CACCAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 4264
QY 850 -----ProSerProSerSerSerProSerProSerProSerProSerProSerS 866
```



```

Db 4265 CGGTCGGGACCTTCGGCGGATCTCGCCGCCACCTCAACCAACCCCACTCACT 4324
QY 866 erSerPro 868
Db 4325 CCTCTCCC 4332

RESULT 13
US-09-252-991A-13773/c
; Sequence 13773, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: ABRUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09-02-18
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13773
; LENGTH: 4131
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13773

Alignment Scores:
Pred. No.: 6,77e-06 Length: 4131
Score: 288.50 Matches: 246
Percent Similarity: 31.64% Conservative: 109
Best Local Similarity: 21.93% Mismatches: 376
Query Match: 5.62% Indels: 392
DB: 4 Gaps: 56

US-09-917-376-1 (1-957) x US-09-252-991A-13773 (1-4131)
QY 37 IleAlaIleThrAlaSerProAlaHisAlaAlaThrThrGlnProTyrThrTrpSer--- 55
Db 4113 GTCACGTGACCGGAGCAGCGCGCGGCAATACACAGCGTCCCGCTACCGAGCGTG 4054
QY 56 -----AsnValAlaIleGlyGlyGlyGlyPheValAspGly 67
Db 4053 GATTCCTCGTCGCGTCGATCCGCGAGTGGATCCGAGCAACGGTTTCGGTGATCAGCGC 3994
QY 68 IleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMet 87
Db 3993 -----ACCGCGAGCGCGGCAACACC 3973
QY 88 TyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsn 107
Db 3972 ATCATCATCCGATGGCAACGGCAACCCGATTTGGCCAGGTCAACCGCGACGCGCGT 3913
QY 108 AsnTrpGlyTyr-----AsnGlyValVal-----SerIleAla 118
Db 3912 AACTGGTCTTCACTCCAGCATCCCGTCGCGGATGGCAGCGTGTCAACGTTGGTGGG 3853
QY 119 AlaAspProIleAsnThrAsnIysValTrpAla-----AlaValGlyMet 133
Db 3852 CGCAGCCCAAGCAATGTCACAGATGCGCGCGCGGTGATCACTGTGTGATGCGTGGCCCG 3793
QY 134 TyrThrAsnSerTrpAspProAsnAspGlyAla---IleLeuArgSerSerAspGlnGly 152
Db 3792 GCGGCGCGGTGATCATCGAGCAACGGCAGCAGCATAGCGGTACCGGAGGCGCGC 3733
QY 153 AlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArgGly 172
Db 3732 GCGAGCGGTATCTCACCAGT-----GGCGGCGCAACCG 3697
QY 173 MetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaPro 192

```

```

Db 3696 ATCGCGCAGCGCCACCGCC----- 3679
QY 193 SerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsn 212
Db 3678 GACGGCAGCGGCAACTGGACGTTCAACCCG----- 3649
QY 213 PheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAsp 232
Db 3648 -----GGCACCCTCGTGGCCAAACGGCACCGTGATCAACGCC----- 3613
QY 233 IleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeuGlyGlnAlaSer 252
Db 3612 -----GTGGCCCGCAGGACCGCGCGCAATACCGAGCGTCCGGCAGC 3571
QY 253 LysThrIlePhe-----ValGlyValAlaAspProAsnAsnProValPhe 267
Db 3570 GTCACCGTCGATGCCATCGCCCGCGCGCGGTGATCAATCCGAGCAACGAGATCGTC 3511
QY 268 TrpSer----- 269
Db 3510 ATCAGCGGTACGGCGGAACCGGGGCCACGGTGATCCTCACCGACGGCAACGCCCG 3451
QY 270 -----ArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaPro--- 283
Db 3450 ATCGGCGAGGTCAACCGCGCAGCGGCAAGTGGGCTTTCACGCGCGCCACCGCGCTG 3391
QY 283 ----- 283
Db 3390 GCCAATGGCAGCGTGATCAATGCGTGGCCCGCAGGACGCGCGCGCAACACAGCATGCC 3331
QY 284 -----ThrGlyPheIleProHisGlyValPheAspProValAsnHis 298
Db 3330 ACCAGCGCCACCGTCGATCGCTGGCGCCAGAGCCCCGGTGATCGATCCGAGCAACGGT 3271
QY 299 ValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyr-----AspGlySerSer 315
Db 3270 AGCGTGATCGCGGTACCGCGCGAGCGGTGTGTGCGCACGGTGATCTCACCGACGGCAACGGC 3211
QY 316 GlyAspValTrpLysPheSerValThr---SerGlyThrTrpThrArgIleSerProVal 334
Db 3210 AACCCGATCGCGCAGGTCAACCGCGATGGCAGCGCAACTGGAGCTTCACCGCGCGCAGC 3151
QY 335 Pro-----SerThrAspThrAlaAsnAspTyrPhe 344
Db 3150 CGGCTCTCAATGGCAGCGTGTCATGCGGTGGCCGCGCGCGCGGTGATCGACCGCAGC 3091
QY 345 GlyTyrSerGlyLeuThrIleAspArgGlnHisPro----- 356
Db 3090 GGCCCGCGCAGCACCGTCGTTGACTCGGTGGCGCGCGCGCGGTGATCGACCGCAGC 3031
QY 357 AsnThrIleMetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSer 376
Db 3030 AACCGCAGCGTGATCGCGGTACCGCGGAAGCGGTGCGAGCGGTGATCTCTC----- 2980
QY 377 ThrAspGlyGly-----AlaThrTrpThrArgIleTyrAspTrpThr 390
Db 2979 ACCGATGGCGCGGCAACCCGATCGCGCGACCGCCAGCCCGATGGCAGCGCAACTGGAGC 2920
QY 391 SerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAla----- 405
Db 2919 TTCACCCCGCGCAGCGCGTGGCCAAACGGCACCGGTGATCAATGCGGTGGCCAGGATCCG 2860
QY 406 -----GluProTrpLeuThrPhe-----GlyValGlnProAsnProPro 418
Db 2859 GCGCGCAATACCGAGCGCGCGCACGACCGACCGTGGACGCGGTGGCGCGCGCCAGCCCG 2800
QY 419 ValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSer 438
Db 2799 GTGGTCAACCGGAGCAACCGC----- 2779
QY 439 AspArgMetLeuTyrGlyThr-----GlyValaThrLeuTyrAlaThrAsnAspLeu 455
Db 2778 ----AGCGTGATCGCGGTACCGCGGAGCGCGCGCACGGTGATCTCTCACC----- 2731

```



```
Qy 456 ThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGlu 475
Db 2730 -----GACGCGCGGCAAC-----CGGATCGCCAGGTACACCGCCGAC 2692
Qy 476 ThrAlaValAsnAspLeuIleSerProSerGlyAlaProLeu----- 490
Db 2691 GGCAGCGGCAACTGGAGTTTCACGCC-----GGCAGCGGTGGCCCAACGGCTCGGTG 2638
Qy 491 IleSerAlaLeuGly---AspLeuGlyGlyPheThrHisAlaAspVal---ThrAlaVal 508
Db 2637 ATCAATCGCTGGCCAGGACCGCGCGCAACACAGCGCGCGCGCAGCACCGCTG 2578
Qy 509 ProSerThrIlePheThrSerProValPhe-----ThrThrGlyThrSerValAsp--- 525
Db 2577 GACTCGGTAGCCCGCGCCACCCCGGTGCTGATCCGAGCAACCGGTGCGGTACAGCGGT 2518
Qy 526 TyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSerGln 545
Db 2517 ACCGCGAGCGCGGCGCAGGTGATCTCACGACGGCGCGCAACCCGATACGGCAG 2458
Qy 546 ProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySer 565
Db 2457 GCC-----ACGCGCATGGCAGCGCAACTGG-----AGCTTC 2425
Qy 566 GluProGlyGlyValThrThrGlyGlyThrVal-----AlaAlaSerAlaAspGlySer 583
Db 2424 ACTCCGGGCACACCGCTGACCAACGACGCGGTGATCAATGCGGTGGCCAGGACCGGCC 2365
Qy 584 ArgPheValTrpAlaPro----- 589
Db 2364 GGCAACACGACGCGTCCGTGACGACCAACAGTGGACGCGGTGGCCCGCCACCCCGGTG 2305
Qy 590 -----GlyAspProGlyGlnProValVal 597
Db 2304 ATCGACCCGCAATGGTGTCAAACTCAGCGGCACCGCGCAACCGCGTCCGGTGTATC 2245
Qy 598 TyrAlaValGlyPheGlyAsn-----SerTrp 606
Db 2244 CTCACCGATGGCAATGGCAACCCGATCGCGCAGACCTCGCGCGGTAGCGGCAACTGG 2185
Qy 607 AlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLys 626
Db 2184 ACCTTCACCGCGGACCGCG----- 2164
Qy 627 ThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPhe 646
Db 2163 -----CTGCCAACGGCACTGGTGTCAACGCC-----GTGCCCCAG 2128
Qy 647 GlnProValAla---AlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAla 665
Db 2127 GACCGCGCGGCAATATACCAGCGGTCCGCGCAGCACCGGTGGATACGTCGCGCGCGCC 2068
Qy 666 ValProGlyLysGluGlyAspLeuTrpLeuAlaAspSerGlyLeuTyrHisSerThr 685
Db 2067 ACGCG-----GTATCAATGCCAGC 2047
Qy 686 AsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPhe 705
Db 2046 AACGCG-----AGCGTGTATCCCGCAGCGCC-----GAGGTCCGC--- 2011
Qy 706 GlyLysSerAlaProGlySerSerTyrProAlaValPheValGlyThrIleGlyGly 725
Db 2010 -----GCCAAAGTGTCTCTCACCGCAGCGCAACCGCAAC 1978
Qy 726 ValThrGlyAlaTyrArgSerAspCysGlyThrThrTrpValLeuIleAsnAspAsp 745
Db 1977 CCGATCGCGCAGACACCGCGCAGCGAGTGGC---AACTGGACCTTCACCCCGCGCAG 1921
Qy 746 GlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArg 765
Db 1920 CCGTGGCCAAAGGTACGGTGTATCAACGCCGTGCGGAGACGCGCG----- 1873
```

```
766 ValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyGlyAlaProSer 785
1872 -----GGCAACGCCAGCGGT---CGGCGC 1852
786 GlySerProSerProSerValSerProSerAlaSerProSerLeuSerProSerProSe 805
1851 AGCAACCGGTGGACTCGGTGGCGCGTCCGCT---CGCTGCTGAGCATCAGCGCGCAG 1795
805 rProSerSerSerProSerPro-----SerProSerProSe 817
1794 GCGCGGTGTGACCGGCACCGCGAGCCGAGCAAGTCAAGTGGTGTATCTGTGTCAACGCG 1735
817 rSerSerProSerProSerProSerProSerProSerProSerProSerProSerArgSe 837
1734 GACACCGCAACCGATCACGCTCACCGTCGACGCGCGCGCAACTTCAGCCCTGCGGTTTC 1675
837 r-----ProSerPr 840
1674 GCGCGCGCTGATCACCGCGCAACTGATCGCGGGGTTGCGGTGAGACCGCGCGCAAC 1615
840 oSerAlaSerProSerProSerProSerProSerProSerProSerProSerProSer---Se 859
1614 GTACGCGCGCGCGCCACCATCAACCGCGCGGACCTGGCGCCGCCACCATCAGCGTGGCG 1555
859 rProSerProThrProSerSerProValSerGlyGlyValLysValGlnTyrLysAs 879
1554 GAAGCGCGCGATCTGTGATCAACCGCGCGGAGATCGGGGACGCG-ATCCAGGTC----- 1501
879 nAsnAspSerAlaProGlyAspAsnGlnIleLysProGlyLeuGlnValValAsnThrGl 899
1500 -----GATGTACGCTCGTCCGACCATCAGTCCGCG----- 1468
899 ySerSerSerValAspLeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGl 919
1467 -----CAGGTGCTCAGCGTCAAGTTC---GCCGGCAGACAGCGCTA 1430
919 ySerSerThrLeuValTyrAsnCysAspTrpAlaAlaIleGlyCysGlyAsnIleArgAl 939
1429 CGAGCGCGAGTGCAGCATACCTCACCGCGCGCACATCGCGCGCGCAACTGACCTT 1370
939 aSerPhe-----GlySerValAsnPro-----AlaThrProThrAl 951
1369 GACCTGACGCTCCCGCGCGCATCGCGCGTTCCTCCGAGGCGCGCTCAGCGTCCACCGC 1310
951 aAsp 952
1309 CGAC 1306
RESULT 14
US-09-410-551B-18
; Sequence 18, Application US/09410551B
; Patent No. 6503737
; GENERAL INFORMATION:
; APPLICANT: KOSAN BIOSCIENCES, Inc.
; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTI, DANIEL
; APPLICANT: WU, KAI
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
; FILE REFERENCE: 30062-20026.00
; CURRENT APPLICATION NUMBER: US/09/410,551B
; CURRENT FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: US 60/139,650
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/123,810
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 60/102,748
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
```



```

; LENGTH: 4571
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic
; OTHER INFORMATION: PKs synthase fragment
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (9)..(4559)
; US-09-410-551B-18

```

## Alignment Scores:

```

Pred. No.: 7, 61e-06 Length: 4571
Score: 288.50 Matches: 227
Percent Similarity: 34.22% Conservative: 121
Best Local Similarity: 22.32% Mismatches: 390
Query Match: 5, 62% Indels: 283
DB: 4 Gaps: 45

```

US-09-917-376-1 (1-957) x US-09-410-551B-18 (1-4571)

```

QY 9 LeuThrMetArgSerArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaVal 28
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1662 CTGACCGAACAACAGACCGGCTGCGCGCTACCTGCGGCGTCCGCCGGCGCGATATA 1721
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 29 AlaAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAlaHisAlaAlaThr 48
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1722 CGGGCTGTGGCATCGACGCTG-----GCGGTGACACGCGTGGT-GTTCGAGCACCGCGC 1774
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 49 ThrGlnProTyr----- 52
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1775 CGTACTCTCTGGAGATGACACCGTCACCGGACCGCGGTGACGACCCCGAGATCGTGT 1834
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 53 -----ThrTyrSerAsnValAlaIleGlyGlyGly-----PheVal 65
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1835 TGCTTCTCCGGCGAGGTGGACAGTGGCTGGGGATGGGCGAGTGCACTCGCGGATTCGTC 1894
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 66 AspGlyIleValPheAsnGluGly-----AlaProGlyIleLeuTyrValArgThrAsp 83
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1895 GGTGTGTTCGCGAGCGATGCGCGAGTGTGCGGCGGCTTTCGCGAGTTCGT---GGA 1951
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 84 IleGlyGlyMetTyrArgTyrAspAlaAlaAsnGlyArgTyrIleProLeuLeuAspTyr 103
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1952 CTGGGATCTGTTCACGGTTCGTGATGATCCGGCGGTGG-TGACACGGGTGTGATG---TGG 2007
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 104 ValGlyTyrAsnAsnTyrGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsn 123
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2008 TCACGCCC-GCTTCTGGCGGATG-----ATGTTTCTCTGGCGCG----- 2048
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 124 ThrAsnLysValTyrAlaAlaValGlyMetTyrThrAsnSer---TrpAspProAsnAsp 142
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2049 -----GTGTGGCAGCGCGCGGTGTGCGCGGATGCGGTGATCGGCCATTCGCAG 2099
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 143 GlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTyrGln----- 156
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2100 GGTGAGATCCCGCAGCTGTGTGGCGGTGCGGTGCTACTACGCGATCCGCCCGGATC 2159
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 157 IleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArg 176
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2160 GTGACCTTGGCGACGCGATGCGCGGGGCTGCGGGGCGCGGCGCGATGGCATCC 2219
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 177 LeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGly 196
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2220 GTGCGCCCTG---CCCGCGCAGGATGTCGAGTCTGGTCGACGCGGGCC----- 2261
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 197 LeuTyrArgSerThrAspSerGlyAlaThrTyrSerGlnMetThrAsnPheProAspVal 216
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2262 ---TGGATCGCGCCCAACAGCGCGCGCTCCACCGTGTATCGCGGACCCCGGAGCG 2318
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 217 GlyThrTyrIleAlaAsnProThrAspThrGly----- 228
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2319 GTCGACCATGTCTCTACCGCTCATGAGGCACAGGGGTGCGGGTGCAGCGGATCACCCTC 2378
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

QY 229 ---TyrGlnSerAspIleGlnGlyValValTyrVal-----AlaPheAspLysSer 244
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2379 GACTATGCTCGCACACCCCGCAGCTCGAGTCTCGCGCAGCACTACTCGACATCACT 2438
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 245 SerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsn 264
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2439 AGCGACAGCAGCTCGCAGACCCCGCTGTCGCGTGGCTGTGACCGTGGACGGCACC-- 2495
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 265 ProValPheTyr-----SerArgAspGlyGlyAlaThrTyrGlnAlaValProGlyAla 282
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2496 -----TGGTTCGACAGCCCGCTGGACGGGGAGTACTGGTACCGGAACCTCGCGTAA 2546
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 283 ProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeuTyrIle 302
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2547 CCGGTGCGTTTCCACCCCGCTGACCGAGTTCAGGCCCGCAGGCCGACACCGTGTTCGTC 2606
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 303 AlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTyrLysPheSer 322
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2607 GAGGTCAGCGCCAGCCCGCTGTTGTCAGGGCGAGGACGACGATGTCGACGGTTGCC 2666
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 323 Val-----ThrSerGlyThrTyrThrArgIleSerProValProSerThrAspThr 339
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2667 AGCTCGCTGTCGACGACGCGCGACGCCACCCGGATG-----CTCACCGCCCTG 2714
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 340 AlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIle 359
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2715 GCACAGGCCAT-----GTCCACGCGCTCACCGTCGAC----- 2747
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 360 MetValAlaThrGlnIleSerTyrTyrProAspThrIleIlePheArgSerThrAspGly 379
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2748 -----TGGCGCGCATCTCTC----- 2762
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 380 GlyAlaThrTyrThrArgIleTyrAspTyrThrSerTyrProAsnArgSerLeuArgTyr 399
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2763 GGCACCCACACACCCGGGTACTGACCTTCACCTACGCCCTCCACACACCGCGGTAC 2822
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 400 ValLeuAspIleSerAlaGluProTyrLeuThrPheGlyValGlnProAsnProProVal 419
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2822 ----- 2822
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 420 ProSerProLysLeuGlyTyrMetAspGluAlaMetAlaIleAspProPheAsnSerAsp 439
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2823 -----TGGCTCGAGTCGCGCAGC-----CCGCGCGCATCCGAC 2855
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 440 -----ArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLys 457
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2856 GCGGGCCACCCCGTGTGGCTCCGGTATCGCCCTCGCCCGG----- 2897
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 458 TrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAla 477
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2898 -----TGGCGCGCGCGGGTTCACGGGT----- 2921
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 478 ValAsnAspLeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeu 497
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2922 -----TCCGTGCGCGCGGTGCGGACCGCGCGGTTCGTCGCGCGAGCTG 2966
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 498 GlyGlyPheThrHisAlaAspVal-ThrAlaValProSerThrIlePheThrSerProVal 517
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2967 GCGTGGCGCGCGCGGACGCGTCACTGCGCCCGCGTCCGAGCGCGGTTCACATCGCTCC 3026
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 517 lPheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAl 537
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3027 GTCCCGCGCGCGCGGCGCATGGCCGACCGACCGTACAGACCT----- 3069
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 537 aGlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 557
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3070 -GGGTTCGAC-GAGCCCGCGGACGACGCGCGCGCGGTTCACCGTCACACCCGCGCGG 3127
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 557 yGlyLysAsnTyrPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAl 577
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3128 CGACGCGCGCGTGCACGCTGCACGCGCGAG---GGGTGCTGCGCGCCCATGGCAGC- 3179
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 577 aAlaSerAlaAspGlySerArgPheValTyrAlaProGlyAspProGlyGlnProValVa 597
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```







QY 84 IleGlyGlyMetTyrArgTyrAspAlaAlaAsnGlyArgTyrPheProLeuLeuAspTyr 103  
Db 1952 CTGGGATCTGTTCAGGTTCTGGATGATCCGGCGGTGG-TGGACCGGTTGATG---TGG 2007  
QY 104 ValGlyTyrAsnAsnTyrGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsn 123  
Db 2008 TCCAGCCC-GCTTCTCGCGCATG-----ATGGTTTCCCTGGCGCG----- 2048  
QY 124 ThrAsnLysValTyrAlaAlaValGlyMetTyrThrAsnSer---TyrAspProAsnAsp 142  
Db 2049 -----GTGTGGCAGCGCGCGGTGTGGCGCGATGCGGTGATCGGCATTTCGCAG 2099  
QY 143 GlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTyrGln----- 156  
Db 2100 GGTGAGATCGCGCATGTGTGGCGGTGCGGTGCATCAGCGGATGCGCGCGCATC 2159  
QY 157 IleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArg 176  
Db 2160 GTGACCTTGGCAGCGCAGCGATCGCGGGCCTGGCGGGCGGGCGCGATGCGATCC 2219  
QY 177 LeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGly 196  
Db 2220 GTGCGCCCTG---CCGCGCAGGATGTCAGCTGTCGACGGCGCC----- 2261  
QY 197 LeuTyrArgSerThrAspSerGlyAlaThrTyrSerGlnMetThrAsnPheProAspVal 216  
Db 2262 ---TGGATCGCGCGCCACACAGCGCGCCCTCCACCGTGTATCGCGGGCACC----- 2318  
QY 217 GlyThrTyrIleAlaAsnProThrAspThrThrGly----- 228  
Db 2319 GTGCAACCATGCTCCTCAGCGCTCATGAGGACAGAGGGGTGCGGTGCGGGATCACCGTTC 2378  
QY 229 ---TyrGlnSerAspIleGlnGlyValValTyrVal-----AlaPheAspLysSer 244  
Db 2379 GACTATGCTCGACACCCCGACGTCGAGCTGATCCGCGAGAACTACTCGACATCACT 2438  
QY 245 SerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsn 264  
Db 2439 AGCGACAGCAGCTCGACACCCCGCTGTCGCGCGTGTGTGACCGGTGTCGACCGTGCACGCGCACC--- 2495  
QY 265 ProValPheTyrP-----SerArgAspGlyGlyAlaThrTyrGlnAlaValProGlyAla 282  
Db 2496 -----TGGGTGACAGCCCGCTGAGCGGGAGTACTGTATCCGGAACTCGGTGAA 2546  
QY 283 ProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeuTyrIle 302  
Db 2547 CCGGTGCGTTTCACCCCGCTCAGCCAGTTGCGAGGCCCGCGCGACACCGTGTGTC 2606  
QY 303 AlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTyrLysPheSer 322  
Db 2607 GAGGTGAGCGCCAGCGCGGTGTGTGCGGCGATGACGACGATGTCTCAGCGTTGCC 2666  
QY 323 Val-----ThrSerGlyThrTyrThrArgIleSerProValProSerThrAspThr 339  
Db 2667 ACGTCTCGTGTGACGACGCGACCGCACCCCGATG-----CTCACCGCCCTG 2714  
QY 340 AlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIle 359  
Db 2715 GCACAGCCCTAT-----GTCCAGCGCGTACCGTCGAC----- 2747  
QY 360 MetValAlaThrGlnIleSerTyrProAspThrIleIlePheArgSerThrAspGly 379  
Db 2748 -----TGGCCCGCATCTCTC----- 2762  
QY 380 GlyAlaThrTyrThrArgIleTyrAspThrTyrSerTyrProAsnArgSerLeuArgTyr 399  
Db 2763 GGCACCAACCAACCGCGGTACTGGACCTTCCGACCTACGCTTCCACACCGCGGTAC 2822  
QY 400 ValLeuAspIleSerAlaGluProTyrLeuThrPheGlyValGlnProAsnProProVal 419  
Db 2822 ----- 2822

QY 420 ProSerProLysLeuGlyTyrMetAspGluAlaMetAlaIleAspProPheAsnSerAsp 439  
Db 2823 -----TGGCTCGAGTCCGACGCG-----CGGCGCGCATCCGAC 2855  
QY 440 -----ArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLys 457  
Db 2856 GCGGCCACCCCGTGTGCGGTATCGCCCTCGCGCGG----- 2897  
QY 458 TrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAla 477  
Db 2898 -----TCCCGCGCGGTGTTCACGGT----- 2921  
QY 478 ValAsnAspLeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeu 497  
Db 2922 -----TCCGTGCGCGGTGTTCACGGT----- 2966  
QY 498 GlyGlyPheThrHisAlaAspVal-ThrAlaValProSerThrIlePheThrSerProVa 517  
Db 2967 GCGCTGCGCGCGCGACGCGGTGCTGCGCGACCGTTCGAGCGGTTCGACATCGCCTCC 3026  
QY 517 IPhThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAl 537  
Db 3027 GTGCCCGCGCGCGCGCATGCGCGACGACCGTACAGACCT----- 3069  
QY 537 aGlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGl 557  
Db 3070 -GGGTGCGAC-GAGCCGCGCGACGCGCGCGCGCGTTCACCGTGACACCGCACCGG 3127  
QY 557 yGlyLysAsnTyrPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAl 577  
Db 3128 CGACGCGCGGTGACGCTGCACGCGAG---GGGTGCTGCGCGCGCGTGCACG----- 3179  
QY 577 aAlaSerAlaAspGlySerArgPheValTyrAlaProGlyAspProGlyGlnProValVa 597  
Db 3180 -GCCCTGCGCGATGCGCGCGACCGCGAGTGGCCCCA-----CGCGC-----GCGGT 3226  
QY 597 lTyrAlaValGlyPheGlyAsnSerTyrAlaAlaSerGlnGlyValProAlaAsnAlaGl 617  
Db 3227 GCGCGGAGCGGCTCGCGGTGTGTGCGCGCGGGGACGAGTCTTCGCGGAGCGCA 3286  
QY 617 nIleArgSerAspArg-----ValAsnPro-----LysThrPheTyrAl 630  
Db 3287 GGTGAGCGACCGCGCGTTCGTGTGTGCGACCGCGACCTGCTGCGCGGTCTTCTCGC 3346  
QY 630 aLeuSerAsnGlyThr----- 635  
Db 3347 GGTGCGCGACGGAAGCG 3406  
QY 636 -----PheTyrArgSerThrAspGlyGlyValThrPheGlnPr 648  
Db 3407 CGCCACCGTACTCGCGCGCTGCTCACC CGCGCGCACCGAGCGCGCGCGCGCGCGCG 3466  
QY 648 oVal---AlaAlaGlyLeuProSerGlyAlaValGlyValMetPheHisAlaValPr 667  
Db 3467 CTTTCGAGCGCGCGCGCTGCGGTACTCACCGCGGAGCGGTGACGCGTTCGCGGAGGTGC 3526  
QY 667 oGlyLysGluGlyAsp-----LeuTyrLeuAlaAlaSe 678  
Db 3527 GTCACCGTCCGCTCGCGAGGAGTCCGACCGCGCTGCGCGGTGAGTGTGCGGTGCGC 3586  
QY 678 rSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTyrSerAlaIleThrGlyValSe 698  
Db 3587 CGAGGCGGTCTAC-----GACGTGACCTGCGCGCGAGGACATGTC-----CTGAT 3631  
QY 698 rSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPh 718  
Db 3632 CACCGCGCGCCACCCCGACCGCGAGGACATACCCACCGCGCGCGCGCGCGCGCAC 3691  
QY 718 eValVal----- 720  
Db 3692 CCGCGTCTGACCGCGCTGCAACACCACTCACCCACCGCGCGCGCGCGCGCGCGCG 3751  
QY 721 -----GlyThrIleGlyGlyValThrGlyAlaTyrArgSerAs 733



```

Db      3752  CACCACACCGACCCCGCGCGCTACCGGCTACCGCGCACCGCGCCAGAACGA 3811
Qy      733  pAspCysGlyThrThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGl 753
Db      3812  ACACCCCGACCGCATCGCGCTCATGAAACCGACACCGCCACACCGCCCTCCCGCTGGC 3871
Qy      753  yGlnAlaIleThrGlyAspHisAlaAsnLeuArgArgValTyr----- 767
Db      3872  CCAACTCGCCACCTCGACACCCCGCTCGGCTCACCACACACCCCTCCACCC 3931
Qy      768  -----IleGlyThrAsnGlyArgGlyIleValTyr-GlyAspIleGlyGlyA 783
Db      3932  CCACCTCACCCCTCCACACCCACCCACCCACCCACCCCGCTCAACCCCGACCA 3991
Qy      783  laProSerGlySerPro---SerProSerValSerProSerAlaSerProSerLeuSerP 802
Db      3992  CGCCATCATCATACCGCGGCTCGGCAACCTCGCGGCATCTCGCGCGCACCTGAA 4051
Qy      802  roSerProSerProSerSerProSerPro----- 812
Db      4052  CCACCCCGACACCTACCTCTCTCCGCGACCCCGACCCCGCGCACCCCGCGCACCA 4111
Qy      813  --SerProSerProSerProSerProSerProSerProSerPro---SerProS 831
Db      4112  CCTCCCTCGACGCTCGGCGACCCCGACCACTCGCCACACCCCTCACCCACATCCCCA 4171
Qy      831  erProSerPro-----SerArgSerProS 839
Db      4172  ACCCTCACCGCATCTTCCACACCGCGCGCACCTCGACGCGCATCTCTCCAGCCCT 4231
Qy      839  erProSerAlaSerProSerProSerProSerProSerPro-----S 853
Db      4232  CACCCCGACCGCTCACACCGCTCTCCACCCCAAGCCAAACCGCGCTGGCACCTGCA 4291
Qy      853  erSerSerProSerProSerProSerProSerProSerProSerProSerPro 868
Db      4292  CCACCTCACCAAAACCAACCCCTCACCCACTTGTCTCTACTCCA 4338

```

Search completed: October 5, 2005, 21:26:53  
Job time : 627.52 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 5, 2005, 08:30:02 ; Search time 1843.53 Seconds  
(without alignments)  
3584.301 Million cell updates/sec

Title: US-09-917-376-1  
Perfect score: 5135  
Sequence: 1 MDRSENRLTWSRRLYSLL.....RASFGSVNPATPTADTYLQX 957

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 7442561 seqs, 3452328358 residues

Total number of hits satisfying chosen parameters: 14885122

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US09917376/runat\_04102005\_164329\_26954/app\_query.fasta\_1.2524  
-DB=Published Applications NA -OPM=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09917376 @CGN 1.1 2178 @runat\_04102005\_164329\_26954  
-NCPU=6 -ICPU=3 -NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FCGAPOP=6 -FCAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

- 1: /cgn2\_6/ptodata/1/pubna/US07\_PUBCOMB.seq:
- 2: /cgn2\_6/ptodata/1/pubna/PCT\_NEW\_PUB.seq:
- 3: /cgn2\_6/ptodata/1/pubna/US05\_NEW\_PUB.seq:
- 4: /cgn2\_6/ptodata/1/pubna/US06\_PUBCOMB.seq:
- 5: /cgn2\_6/ptodata/1/pubna/US07\_NEW\_PUB.seq:
- 6: /cgn2\_6/ptodata/1/pubna/PCTUS\_PUBCOMB.seq:
- 7: /cgn2\_6/ptodata/1/pubna/US08\_NEW\_PUB.seq:
- 8: /cgn2\_6/ptodata/1/pubna/US08\_PUBCOMB.seq:
- 9: /cgn2\_6/ptodata/1/pubna/US09A\_PUBCOMB.seq:
- 10: /cgn2\_6/ptodata/1/pubna/US09B\_PUBCOMB.seq:
- 11: /cgn2\_6/ptodata/1/pubna/US09C\_PUBCOMB.seq:
- 12: /cgn2\_6/ptodata/1/pubna/US09\_NEW\_PUB.seq:
- 13: /cgn2\_6/ptodata/1/pubna/US10A\_PUBCOMB.seq:
- 14: /cgn2\_6/ptodata/1/pubna/US10B\_PUBCOMB.seq:
- 15: /cgn2\_6/ptodata/1/pubna/US10C\_PUBCOMB.seq:
- 16: /cgn2\_6/ptodata/1/pubna/US10D\_PUBCOMB.seq:
- 17: /cgn2\_6/ptodata/1/pubna/US10E\_PUBCOMB.seq:
- 18: /cgn2\_6/ptodata/1/pubna/US10F\_PUBCOMB.seq:
- 19: /cgn2\_6/ptodata/1/pubna/US10G\_PUBCOMB.seq:
- 20: /cgn2\_6/ptodata/1/pubna/US10H\_PUBCOMB.seq:
- 21: /cgn2\_6/ptodata/1/pubna/US10I\_PUBCOMB.seq:
- 22: /cgn2\_6/ptodata/1/pubna/US10J\_NEW\_PUB.seq:
- 23: /cgn2\_6/ptodata/1/pubna/US11A\_PUBCOMB.seq:
- 24: /cgn2\_6/ptodata/1/pubna/US11\_NEW\_PUB.seq:
- 25: /cgn2\_6/ptodata/1/pubna/US60\_NEW\_PUB.seq:
- 26: /cgn2\_6/ptodata/1/pubna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	5134	100.0	2869	11	US-09-917-376-2	Sequence 2, Appli
2	5134	100.0	2869	15	US-10-155-400-2	Sequence 2, Appli
3	2487	48.4	9025608	15	US-10-156-761-1	Sequence 1, Appli
4	2478	48.3	2646	15	US-10-156-761-1845	Sequence 1845, Ap
5	1631.5	31.8	2517	15	US-10-026-994-4	Sequence 4, Appli
6	1631.5	31.8	2710	15	US-10-026-994-1	Sequence 1, Appli
c	1570	30.6	9025608	15	US-10-156-761-1	Sequence 1, Appli
8	1442	28.1	2217	15	US-10-156-761-2561	Sequence 2561, Ap
9	1167.5	22.9	3668	10	US-09-927-827-21	Sequence 21, Appl
10	1125.5	21.9	2646	18	US-10-395-241-11	Sequence 11, Appl
11	1124	21.9	2481	18	US-10-395-241-17	Sequence 17, Appl
12	1117	21.8	2367	18	US-10-395-241-13	Sequence 13, Appl
13	940	18.3	5698	18	US-10-420-191-1	Sequence 1, Appli
14	823	16.0	3687	10	US-09-917-384-2	Sequence 2, Appli
15	823	16.0	3687	10	US-09-917-384-2	Sequence 2, Appli
16	702	13.7	2289	10	US-09-917-378-2	Sequence 2, Appli
17	392.5	7.6	1103	20	US-10-653-047-7511	Sequence 7511, Ap
c	365.5	7.1	651	22	US-10-450-763-20760	Sequence 20760, A
19	361	7.0	2223	15	US-10-156-761-550	Sequence 550, App
20	346.5	6.7	2706	21	US-10-933-262-11	Sequence 11, Appl
c	345	6.7	18596	9	US-09-880-107-1590	Sequence 1590, Ap
c	345	6.7	18596	9	US-09-967-768A-119	Sequence 119, App
c	345	6.7	18596	9	US-09-954-531-124	Sequence 124, App
c	345	6.7	18596	9	US-09-954-531-348	Sequence 348, App
c	345	6.7	18596	19	US-10-629-313-140	Sequence 140, App
c	345	6.7	18596	21	US-10-843-641A-1191	Sequence 1191, Ap
c	345	6.7	18596	21	US-10-843-641A-1415	Sequence 1415, Ap
c	345	6.7	18596	21	US-10-843-641A-6264	Sequence 6264, Ap
c	331	6.4	1887	21	US-10-939-262-13	Sequence 13, Appl
30	330.5	6.4	7407	16	US-10-246-330-3	Sequence 3, Appli
31	330.5	6.4	7407	17	US-10-282-122A-30151	Sequence 30151, A
32	319.5	6.2	9495	19	US-10-669-161-71	Sequence 71, Appl
33	299.5	5.8	210528	13	US-10-087-192-289	Sequence 289, App
34	297	5.8	32591	17	US-10-085-117-187	Sequence 187, App
35	296.5	5.8	4767	10	US-09-940-316B-28	Sequence 28, Appl
36	296	5.8	96602	17	US-10-085-117-61	Sequence 61, Appl
37	293	5.7	4818	10	US-09-940-316B-32	Sequence 32, Appl
c	292.5	5.7	185548	14	US-10-175-523-62	Sequence 62, Appl
c	292.5	5.7	185548	24	US-11-099-265-62	Sequence 62, Appl
c	290	5.6	1232	19	US-10-761-169-17	Sequence 17, Appl
c	289	5.6	324	17	US-10-282-122A-13296	Sequence 13296, A
42	288.5	5.6	4571	10	US-09-940-316B-18	Sequence 18, Appl
43	288	5.6	288	17	US-10-437-708-111	Sequence 111, App
44	288	5.6	288	17	US-10-395-402-111	Sequence 111, App
45	288	5.6	288	20	US-10-257-199-111	Sequence 111, App

ALIGNMENTS

RESULT 1  
US-09-917-376-2  
; Sequence 2, Application US/09917376  
; Publication No. US2004003834A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: 40197.4US01  
; CURRENT APPLICATION NUMBER: US/09/917,376  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1



; SEQ ID NO 2  
; LENGTH: 2869  
; TYPE: DNA  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; NAME/KEY: modified\_base  
; LOCATION: (2869)  
; OTHER INFORMATION: a, c, t, g, other or unknown  
US-09-917-376-2

Alignment Scores:  
Pred. No.: 0 Length: 2869  
Score: 5134.00 Matches: 956  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 99.98% Indels: 0  
DB: 11 Gaps: 0

US-09-917-376-1 (1-957) x US-09-917-376-2 (1-2869)

Qy 1 MetAspArgSerGluAsnIleArgLeuThrMetArgSerArgArgLeuValSerLeuLeu 20  
Db 1 ATGGATCGTTCGGAGAACATCCGCTGACTATGATGATCAGACGATTGGTATCACTGCTC 60  
Qy 21 AlaAlaThrAlaSerPheAlaValAlaAlaAlaLeuGlyValLeuProIleAlaIleThr 40  
Db 61 GCCGCCACTGGCTGTTCCGCGTGGCCGCCCTCTGGGAGTTCTGCCCATCGCGATAACG 120  
Qy 41 AlaSerProAlaHisAlaAlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGly 60  
Db 121 GCTTCTCTCGCGACGGCGGACGACTCAGCCGTACACCTCGAGCAACGCTGGCGATCGG 180  
Qy 61 GlyGlyGlyPheValAspGlyIleValPheAsnGlnGlyAlaProGlyIleLeuTyrVal 80  
Db 181 GCGCGCGCTTGTTCGACGGATCGTCTTCAATGAAGTGCACCGGGAATCTGTACGTG 240  
Qy 81 ArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeu 100  
Db 241 CGGACGGACATCGGGGGATGTATCGATGGGATGCCGCAACGGCGGTGATCCCTCTT 300  
Qy 101 LeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAsp 120  
Db 301 CTGGATTGGGTGGATGGAACAATTTGGGGGTACAAACGCGCTCGTCAGCATTTGGCGAC 360  
Qy 121 ProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspPro 140  
Db 361 CCGATCAATACTAAACAGGTATGGCGCCGCTCGGAATGTACACAAACAGCTGGGACCCA 420  
Qy 141 AsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeu 160  
Db 421 AACGACGGAGCGATTCTCCGCTCGTCTGATCAGGGCGCAACGTGGCAAAATTAACGCCCTG 480  
Qy 161 ProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAsp 180  
Db 481 CCGTTCAGCTTGGCGGCAACATGCCGGGGTGGAAATGGGGAGCGGCTTGGGTGGAT 540  
Qy 181 ProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSer 200  
Db 541 CCAACAATGACAAATTTCTGATTTCGGCGCCCGCGAGCGGCAAAAGGCTCTGGAGAAAGC 600  
Qy 201 ThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIle 220  
Db 601 ACAGATTCCGGCGGACCTTGTCTCCAGATGACGAATTTCCGGACGTAGGCACGTACATT 660  
Qy 221 AlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAla 240  
Db 661 GCAAAATCCACTGACACACCGCGTATCAGAGCGATATTCAGAGGCTCGTCTGGGTCCGT 720  
Qy 241 PheAspLysSerSerSerLeuGlnAlaSerLysThrIlePheValGlyValAla 260  
Db 721 TTCGACAGTCTTCGTCTATCGTCCGGCAAGCGAGTAAGACCAATTTTGTGGCGGTGGCG 780  
Qy 261 AspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValPro 280

Db 781 GATCCCAATAATCCGCTCTTCTGGAGCAGACGCGCGCGACGTGGCAGCGGTGCGG 840  
Qy 281 GlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeu 300  
Db 841 GGTGCGCGACCGCGCTTCATCCCGCACAGGGCGCTTTTGACCCCGTCAACACACGTGCTC 900  
Qy 301 TyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLys 320  
Db 901 TATATTGCCACACGCAATACGGGTGTCGTTATGACGGGAGCTCCGGCGACGCTCTGAAA 960  
Qy 321 PheSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAspThrAla 340  
Db 961 TTTCTCGGTGACTCCGGGACATGGACGCGAATCAGCCCGGTACCTTCAGCGGACACGGCC 1020  
Qy 341 AsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMet 360  
Db 1021 AACGACTACTTTGGTTACAGCGGCTCACTATCGACCGCCAGCACCCGCAACAGATAATG 1080  
Qy 361 ValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGly 380  
Db 1081 GTGGCAACCCAGATATCTGTGGCGGACACCATATCTTTGGAGCACCGACGGCGT 1140  
Qy 381 AlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrVal 400  
Db 1141 GCGACGTGGACGCGGATCTGGGATTGGACGAGTATCCCAATCGAAGCTTGGCATATGTG 1200  
Qy 401 LeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProProValPro 420  
Db 1201 CTTGACATTTCCGGCGAGCCTTTGGCTGACCTTCGCGGTACACCGCAATCCTCCGTACCC 1260  
Qy 421 SerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg 440  
Db 1261 AGTCCGAAGCTCGCTGGATGGATGAGCGCATGGCAATCGATCCGTTCAACTCTGATCG 1320  
Qy 441 MetLeuTyrGlyThrGlyValaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSer 460  
Db 1321 ATGCTCTACGGAACACGCGCGACGTTGTACGCAACAATGATCTCACGAAGTGGGACTCC 1380  
Qy 461 GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAsp 480  
Db 1381 GCGCGCCAGATTCATATCGCGCGGATGGTCAAAAGGATTTGGAGGAGACGCGCGTAAACGAT 1440  
Qy 481 LeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPhe 500  
Db 1441 CTCATCAGCCCGCGCTGTGGCGCCCGCTCATCAGCGCTCTCGAGACCTTCGGCGGCTTC 1500  
Qy 501 ThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThr 520  
Db 1501 ACCCAGCGCGAGTTACTGCGGTGCCATCGACGATCTTCACGTCACCGGTGTTACGACC 1560  
Qy 521 GlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPhe 540  
Db 1561 GGCACACGCTCGCATATCGGAAATGAAATCGCTCGATCATCGTTCGCTCGCGCTGGAAGTTTC 1620  
Qy 541 AspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsn 560  
Db 1621 GATCATCGAGCAACCCGCAACGACGAGCGCTCGCGCTTCGACACGCGCGGCAAGAAC 1680  
Qy 561 TrpPheGlnGlySerGluProGlyValThrThrGlyGlyThrValAlaAlaSerAla 580  
Db 1681 TGGTTCAGGCGCAGCAACCTGCGGGGTGACAGCGGGCGGCACCGTCCGCGCATCGGCC 1740  
Qy 581 AspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaVal 600  
Db 1741 GACGGCTCTCGTTCTGCTGGGCTCCCGCGCATCCCGGTCAAGCTTGTGGTGTACGCGATC 1800  
Qy 601 GlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSer 620  
Db 1801 GGATTTGGCAACTCTCTGGGCTCTTCGCAAGGTGTTCGCGCAATGCCAGATCCCGCTCA 1860  
Qy 621 AspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThr 640



```

1861 GACCGGGTGAATCCAAAGACTTTCTATGCCCTATCAATGGAACCTTCTATCGAAGCAGC 1920
641 AspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGly 660
1921 GACGCGCGCGTGACATTCACACCGGTCCGCGCGGTCTTCCGAGCAGCGGTCCGCGTGGT 1980
661 ValMetPheHisAlaValProGlyLysGluGlyAspLeuTyrLeuAlaAlaSerSerGly 680
1981 GTGATGTTCCACCGCGGTCCCTGGAAAGAGCGGATCTGTGGCTCGCTGCATCGACCGGG 2040
681 LeuTyrHisSerThrAsnGlyGlySerSerTyrPheAlaIleThrGlyValSerSerAla 700
2041 CTTTACCACTCAACCAATGGCGGACGAGTGTCTGCAATCACCGCGGTATCTCCGCG 2100
701 ValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValVal 720
2101 GTGAACGTGGGATTTGGTAAGTCTGCGCGCGGTCTGATACCCAGCGGTCTTTGCTGTC 2160
721 GlyThrIleGlyGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpVal 740
2161 GGCACGATCGGAGCGGTTACGCGGCGGTACCGCTCCGACGACTGGGACGACCTGGGTA 2220
741 LeuIleAsnAspAspGlnHisGlnTyrGlyAsnTyrGlyGlnAlaIleThrGlyAspHis 760
2221 CTGATCAATGATGACGACCAATACGGAATTTGGGACAGCAATCACCGGTGACCCAC 2280
761 AlaAsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyValValTyrGlyAspIle 780
2281 GCGAATTTAGCGCGGTGTACATAGGACGACGACGCGGTGGAAATGTATACGGGACATT 2340
781 GlyAlaProSerGlySerProSerProSerProSerProSerProSerProSerProSer 800
2341 GGTGTGGCGCGGTCCGCGGTCCGCGTCCGCGGTGAGTCCGCTCGGCTTCGCGCGACCTG 2400
801 SerProSerProSerProSerProSerProSerProSerProSerProSerProSerPro 820
2401 AGCCGAGCCCGAGCCCGAGCGAGCTCGCCATCGCCGTCCGCTCGGCTTCGCGCGACCTG 2460
821 SerSerSerProSerProSerProSerProSerProSerProSerProSerProSerPro 840
2461 TCCTCGTCCGCGTCTCCGCGCGTCCGCGTCCGCGTCCGCTTCGCTTCGCTCCGTCACCA 2520
841 SerAlaSerProSerProSerProSerProSerProSerProSerProSerProSerPro 860
2521 TCGCGGTCCGCGCGCGTCTTCGTCACCGAGCCCGTCTTCGTCACCGCTTCGCTCCGCG 2580
861 SerProThrProSerProSerProSerProSerProSerProSerProSerProSerPro 880
2581 AGCCCAACCGCGTCTCGTCCGCGGTGTCGGGTGGGTTGAGGTGCAATGATAGAAATAT 2640
881 AspSerAlaProGlyAspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySer 900
2641 GATTCCGCGCGCGGTGATTAATCAGATCAAGCCGGTTTTCAGGTGGTGAATACCGGTCG 2700
901 SerSerValAspLeuSerThrValThrValArgTyrTrpPheThrArgAspGlyLys 920
2701 TCGTGGTGGATTTGTCGACGCGTGCAGCGTGGTACTGGTTTACCCGGGATGTCGCTCG 2760
921 SerThrLeuValTyrAsnCysAspTyrAlaAlaIleGlyCysGlyAsnIleArgAlaSer 940
2761 TCACACTGGTGTACAACTGTGACTGGGCGGCGATCGGGTGGGAATATCCGCGCTCG 2820
941 PheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 956
2821 TTCGCTCGGTGACCCGCGAGCGCGACCGCGGACACTACTCTGACG 2868

```

RESULT 2

```

US-10-155-400-2
; Sequence 2, Application US/10155400
; Publication No. US2003010898A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.

```

```

; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOYTICUS
; FILE REFERENCE: NREL 01-36A
; CURRENT APPLICATION NUMBER: US/10/155,400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2869
; TYPE: DNA
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (2869)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-10-155-400-2

```

```

Alignment Scores:
Pred. No.: 0 Length: 2869
Score: 5134.00 Matches: 956
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.98% Indels: 0
DB: 15 Gaps: 0

```

US-09-917-376-1 (1-957) x US-10-155-400-2 (1-2869)

```

QY 1 MetAspArgSerGluAsnIleArgLeuThrMetArgSerArgArgLeuValSerLeuLeu 20
Db 1 ATGGATCGTTCGAGAACATCCGTCGTGACTATGAGATCAGACGATTTGGTATCAGTCGTC 60
QY 21 AlaAlaThrAlaSerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThr 40
Db 61 GCGGCCACTCGCTCGTTCGCGCGGTCCGCGCTCTCGGAGTTCTGCGCATCGCGATAACG 120
QY 41 AlaSerProAlaHisAlaIleThrGlnProTyrThrTrpSerAsnValAlaIleGly 60
Db 121 GCTTCTCTCCGCGACGCGCGGACGACTCAGCCGCTACCTGGAGCAACGTTGGCGATCGGG 180
QY 61 GlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrVal 80
Db 181 GCGCGCGGTTCGTCACCGGATCGTCTTCATGAAGGTGACCGGAAATCTGTACGTG 240
QY 81 ArgThrAspIleGlyMetTyrArgTyrAspAlaAlaAsnGlyArgTyrIleProLeu 100
Db 241 CGGACGACATCGGGGGATGATCGATGGATGCCCAACGCGCGGTGGATCCCTCTT 300
QY 101 LeuAspTyrValGlyTrpAsnAsnTyrGlyTyrAsnGlyValValSerIleAlaAlaAsp 120
Db 301 CTGGATGGGTGGATGGAAACAATTCGGGGGTCAACGCGCGTCTGTCAGCATTCGCGGACAG 360
QY 121 ProIleAsnThrAsnLysValTyrAlaAlaValGlyMetTyrThrAsnSerTyrAspPro 140
Db 361 CCGATCAATACTAACAGGTATGGCGCGCTCGGATGTACACCAACAGCTGGGACCCA 420
QY 141 AsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeu 160
Db 421 AACGACGAGCGGATTCCTCGCTCGTATCAGCGCGCAACGTTGGCAATAAATACGCCCTG 480
QY 161 ProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAsp 180
Db 481 CCGTTCAAGCTTGGCGGCAACATSCCGCGCGTGGAAATGGCGGAGCGGCTTGGCGTGGAT 540
QY 181 ProAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSer 200
Db 541 CCAAAACAATGACAAATTCGTATTTTCGGCGCGCCCGGAGCGGCAAGGGCTCTGGAGAACG 600
QY 201 ThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIle 220
Db 601 ACAGATTCGCGCGGACCTGGTCCAGATGACGAACCTTTCCGAGCGTAGGACAGCTACATT 660

```



Qy 221 AlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrrpValala 240  
Db |||||  
Qy 661 GCAAATCCCACTGACAGACCGGCTATCAGAGCGATATTCAAGGCGTGGTGGTGGT 720  
Db |||||  
Qy 241 PheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValala 260  
Db |||||  
Qy 721 TCGACAAAGTCTTCGTCTCATCGCTCGGCGAAGCGAGTAAGACCAATTTTGTGGCGTGGCG 780  
Db |||||  
Qy 261 AspProAsnProValPheTrrpSerArgAspGlyGlyAlaThrTrrpGlnAlaValPro 280  
Db |||||  
Qy 781 GATCCCCAATAATCCGGTCTTCTGGAGCAGAGACGGCGCGGACGTGGAGCGGTGGCG 840  
Db |||||  
Qy 281 GlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeu 300  
Db |||||  
Qy 841 GGTGGCGCGAGCGGCTTCATCCCGCACAAAGGGCGTCTTTGACCGCGTCAACCACTGCTC 900  
Db |||||  
Qy 301 TyrIleAlaThrSerAsnThrGlyProTrrpTrrpAspGlySerSerGlyAspValTrrpLys 320  
Db |||||  
Qy 901 TATATTGCCACCAAGCAATACGGGTGGTCCGTATGACGGAGCTCCGGCGAGCTCGGAAA 960  
Db |||||  
Qy 321 PheSerValThrSerGlyThrTrrpThrArgIleSerProValProSerThrAspThrAla 340  
Db |||||  
Qy 961 TTCTCGGTGACCTCCGGGACATGACCGGAATCATCGCCGCTGACCTTCGACGGACACGGCC 1020  
Db |||||  
Qy 341 AsnAspTrrpPheGlyTrrpSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMet 360  
Db |||||  
Qy 1021 AACGACTACTTTGGTTACGGGCTCACTATCGACCGCCGACACCGCAACGATTAATG 1080  
Db |||||  
Qy 361 ValAlaThrGlnIleSerTrrpTrrpProAspThrIleIlePheArgSerThrAspGlyGly 380  
Db |||||  
Qy 1081 GTGGCAACCCAGATATCGTGGTGGCGGACACCAATAATCTTTTCGGAGCACCGCGGGT 1140  
Db |||||  
Qy 381 AlaThrTrrpThrArgIleTrrpAspTrrpThrSerTrrpProAsnArgSerLeuArgTrrpVal 400  
Db |||||  
Qy 1141 GCGAGTGGAGCGCGATCTGGGATTTGGACGAGTATCCCAATCGAAGCTTCGGGATATGTG 1200  
Db |||||  
Qy 401 LeuAspIleSerAlaGluProTrrpLeuThrPheGlyValGlnProAsnProValPro 420  
Db |||||  
Qy 1201 CTTGACATTTTCGGCGGAGCCTTGGCTGACCTTCGGCGTACAGCGAATCTCCCGTACCC 1260  
Db |||||  
Qy 421 SerProLysLeuGlyTrrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg 440  
Db |||||  
Qy 1261 AGTCCGAAGCTCGGCTGGATGATGAAGCGATGCAATCGATCGCTTCAACTCTGATCGG 1320  
Db |||||  
Qy 441 MetLeuTrrpGlyThrGlyAlaThrLeuTrrpAlaThrAsnAspLeuThrLysTrrpAspSer 460  
Db |||||  
Qy 1321 ATGCTCTACGGAAACAGGCGCGACGTGTGTACGAAACAAATGATCTCACGAAGTGGGACTCC 1380  
Db |||||  
Qy 461 GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAsp 480  
Db |||||  
Qy 1381 GCGCGGACAGATTCATATCGCGCCGATGTTGTCAAAGGATTTGGAGAGACGGCGGTAAACGAT 1440  
Db |||||  
Qy 481 LeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPhe 500  
Db |||||  
Qy 1441 CTGATCAGCGCGCGCTGGCGCCCGCTCATCAGCGCTCTCGGAGACCTCGGCGGCTTC 1500  
Db |||||  
Qy 501 ThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThr 520  
Db |||||  
Qy 1501 ACCCACCGCAGCTTACTCGCGTGCATTCGACGATCTTCAGCTCACCGGTGTTACAGACC 1560  
Db |||||  
Qy 521 GlyThrSerValAspTrrpAlaGluLeuAsnProSerIleIleValArgAlaGlySerPhe 540  
Db |||||  
Qy 1561 GGCACCAAGCTGACTATGCGGAATGGAATCCGTGATCATCTGTCGGCTGGAGATTTTC 1620  
Db |||||  
Qy 541 AspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsn 560  
Db |||||  
Qy 1621 GATCCATCGACCAACCGAAGACAGACGACGCTCGGCTTCGACAGACGGCGGCAAGAAC 1680  
Db |||||  
Qy 561 TrrpPheGlnGlySerGluProGlyGlyValThrThrGlyIleThrValAlaAlaSerAla 580  
Db |||||  
Qy 1681 TGGTTCCAAAGGACGCAACCTGGCGGGTGGAGCGGCGGACCGTCCGCGCATCGGCC 1740  
Db |||||  
Qy 581 AspGlySerArgPheValTrrpAlaProGlyAspProGlyGlnProValValTrrpAlaVal 600  
Db |||||

Db 1741 GACGGCTCTCGTTCGTCTGGGCTCCCGCGCATCCCGGTACGCTGTGTGTACGCAGTC 1800  
Qy GlyPheGlyAsnSerTrrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSer 620  
Db |||||  
Qy 1801 GGATTTGGCAACTCTCGGGCTGCTTCGCAAGGTGTTCCCGCAATGCCAGATCCGCTCA 1860  
Db |||||  
Qy 621 AspArgValAsnProLysThrPheTrrpAlaLeuSerAsnGlyThrPheTrrpArgSerThr 640  
Db |||||  
Qy 1861 GACCGGTGATCCAAAGACTTTCTATGCCCTATCCAATGGAACCTTCTATCGAAGCACG 1920  
Db |||||  
Qy 641 AspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGly 660  
Db |||||  
Qy 1921 GACGGCGGTGATCCAAACCGGTCCGCGCGGTCTTCCGAGCAGCGGTGCGTGGT 1980  
Db |||||  
Qy 661 ValMetPheHisAlaValProGlyLysGlyAspLeuTrrpLeuAlaAlaSerSerGly 680  
Db |||||  
Qy 1981 GTCATGTTCCACGGCGGTGCTGGAAAAGAGCGATCTGTGGCTCGCTCATCGAGCGGG 2040  
Db |||||  
Qy 681 LeuTrrpHisSerThrAsnGlyLysSerTrrpSerTrrpSerAlaIleThrGlyValSerSerAla 700  
Db |||||  
Qy 2041 CTTTACCACCTCAACCAATGGCGGACGAGTGGTCTGCAATCACCGCGGTATCTCCGGG 2100  
Db |||||  
Qy 701 ValAsnValGlyPheGlyLysSerAlaProGlySerSerTrrpProAlaValPheValVal 720  
Db |||||  
Qy 2101 GTGAACGTGGGATTTGGTAACTCTCGCGCGCGGTGCTCATACCCAGCCGTCTTTGTGCTC 2160  
Db |||||  
Qy 721 GlyThrIleGlyGlyValThrGlyAlaTrrpArgSerAspAspCysGlyThrThrTrrpVal 740  
Db |||||  
Qy 2161 GGCACGATCGAGGGGCTTACGGGGCGGTACCGCTCCGACGCTGTGGAGACGCTCGGTA 2220  
Db |||||  
Qy 741 LeuIleAsnAspAspGlnHisGlnTrrpGlyAsnTrrpGlyGlnAlaIleThrGlyAspHis 760  
Db |||||  
Qy 2221 CTGATCATGATGACACACCAATACGGAATTTGGGACACAGCAATCACCGGTGACCAC 2280  
Db |||||  
Qy 761 AlaAsnLeuArgArgValTrrpIleGlyThrAsnGlyArgGlyIleValTrrpGlyAspIle 780  
Db |||||  
Qy 2281 GCGAATTTACGGCGGTGTACATAGGCACGAACGCGCGTGAATTTATATACGGGGACATT 2340  
Db |||||  
Qy 781 GlyGlyAlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeu 800  
Db |||||  
Qy 2341 GGTGTGTGGCGCTCCGGATCGCGCTCTCCGTGCGTGAGTCCGTGGCTTCGCGAGCCCTG 2400  
Db |||||  
Qy 801 SerProSerProSerProSerProSerProSerProSerProSerProSerProSerPro 820  
Db |||||  
Qy 2401 AGCCGAGCCCGAGCCGAGACGCTCGCCATCGCGTCCGCTCGCCGCGAGCTCGAGTCCA 2460  
Db |||||  
Qy 821 SerSerSerProSerProSerProSerProSerProSerProSerProSerProSerPro 840  
Db |||||  
Qy 2461 TCCTCGTCCGCTCTCCGTGCGCGTCACCATCGCGAGTCCGTCTCGTCTCGTCCACCA 2520  
Db |||||  
Qy 841 SerAlaSerProSerProSerProSerProSerProSerProSerProSerProSerPro 860  
Db |||||  
Qy 2521 TCGGCGTCCGCGAGCCGCTCTTCGTACCGAGCCGCTCTTCGTACCCGCTTCGTCGCGCG 2580  
Db |||||  
Qy 861 SerProThrProSerSerProValSerGlyValLysValGlnTrrpLysAsnAsn 880  
Db |||||  
Qy 2581 AGCCCAAACGCGCTCTGTCGCGGTGTCGGGTGGGTGAAGTGCAGTATAGAATAAT 2640  
Db |||||  
Qy 881 AspSerAlaProGlyAspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySer 900  
Db |||||  
Qy 2641 GATTCGGCGCGGTGATAATCAGATCAAGCCGGGTTTCGAGGTGTGAATACCGGGTCC 2700  
Db |||||  
Qy 901 SerSerValAspLeuSerThrValThrValArgTrrpTrrpPheThrArgAspGlyGlySer 920  
Db |||||  
Qy 2701 TCGTGGTGGATTTGTCGACGCTGACGCTGCGGTACTGCTGTTACCCCGGATGTGGTGGT 2760  
Db |||||  
Qy 921 SerThrLeuValTrrpAsnCysAspTrrpAlaAlaIleGlyCysGlyAsnIleArgAlaSer 940  
Db |||||  
Qy 2761 TCGACACTGGTGTACAACTGTACCTGGCGGCGGATCGGGTGTGGGAATATCCCGGCTCG 2820  
Db |||||  
Qy 941 PheGlySerValAsnProAlaThrProThrAlaAspThrTrrpLeuGln 956  
Db |||||







QY 526 TyrAlaGluLeuAenProSerIleIleValArgAlaGlySerPheAspProSerSerGln 545  
 Db 2278924 TACCGCGAGACCGCCCGGACCGTGTGCGGTGCGCAATCTCGAC-----TCGGGT 2278977  
 QY 546 ProAenAspArgHisValAlaPheSerThrAspGlyGlyLysAenTrpPheGlnGlySer 565  
 Db 2278978 CCG-----CATGTGCGGTCTTCGACGCGACAAAGCGCGCAACTGTGTTCGCGGGCGG 2279028  
 QY 566 GluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPhe 585  
 Db 2279029 GACCTTTCGGGGGTTCAGCGGGGTGGAGCGGTTCGCGCGGTTCGACGCGAGTGCCTTC 2279088  
 QY 586 ValTrpAlaProGlyAspProGlyGlnProValValTyrAlaValAlaGlyPheGlyAenSer 605  
 Db 2279089 GTGTGGAGCCCGCGGCGGACCGG-----GTGCAGTACACGACCGGGTTCGACCTCG 2279142  
 QY 606 TrpAlaAlaSerGlnGlyValProAlaAenAlaGlnIleArgSerAspArgValAenPro 625  
 Db 2279143 TGGTCGGGTTCGCGGGCTCCCGCGCGGGGATCGTTCGAGTCCGACCGGTTCGACCGG 2279202  
 QY 626 LysThrPheTyrAlaLeuSerAenGlyThrPheTyrArgSerThrAspGlyGlyValThr 645  
 Db 2279203 AAGACCTTCTACGGCTTCAAGTCCGCGCAGGTTCACGTCTGAGTTCGACGCGGGCGGAC 2279262  
 QY 646 PheGlnProValAlaAala---GlyLeuProSerSerGlyAlaValGlyValMetPheHis 664  
 Db 2279263 TTCACGGGTTCGCGGGCGGCGGCTTCGCGAGCGCGACAGC-----GTGCGCTTCAAG 2279316  
 QY 665 AlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSer----- 679  
 Db 2279317 CGCTGCGCGGACGACGAGCGGACATCTGGTGGCGGCGCGGCGGCGGCGGTAC 2279376  
 QY 680 GlyLeuTyrHisSerThrAenGlySerSerTrpSerAlaIleThrGlyValSerSer 699  
 Db 2279377 GGGTGTGGCACTCGACGCGCGCGCGCGCTTCCACCAAGTCCGCCACCGTTCGACCG 2279436  
 QY 700 AlaValAenValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheVal 719  
 Db 2279437 GCGGACACCATCGGCTTCGCGAAGCGCGGCGCGCGCTCGTACCGACGCTCTACACC 2279496  
 QY 720 ValGlyThrIleGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrTrp 739  
 Db 2279497 AGCGGAGATCGCGGTGTGGCGGCGATCTTCGCTCGCGACGAGCGCGGCGGTGG 2279556  
 QY 740 ValLeuIleAenAspAspGlnHisGlnTyrGlyAenTrpGlyGlnAlaIleThrGlyAsp 759  
 Db 2279557 ACCCGCGTCAACGACGATGCCACCGTGGGTTCGACGCGGCGCGCGATCACCGGTGAC 2279616  
 QY 760 HisAlaAenLeuArgArgValTyrIleGlyThrAenGlyArgGlyValThrGlyAsp 779  
 Db 2279617 CCCAGGGTCTACGGCGCGGTGTATGTGTGTCGACGACGCGCGCGGTATCGTCTACCGCGAC 2279676  
 QY 780 IleGlyGlyAlaProSerGly-----SerProSerProSerValSerProSerAla 796  
 Db 2279677 ACCCGGGCTCTTCGAGCGCGCGGTACGAGCGCGCGCGCGCGCGCGCGCGGTG 2279736  
 QY 797 SerProSerLeuSer-ProSerProSerProSerSerSerProSerProSerProSer 815  
 Db 2279737 ACATACAGGATCACGACGAGTGGTGGCGCGCTTCCAGGCGGATGATACAGTCCGCCAAC 2279796  
 QY 815 rProSerSerSerProSerSerSer-----Pr 824  
 Db 2279797 ACCGGGTTCGACCGCGCTGGAGCGGCTGCTGCTGGGCTGCTGCTGCGTACGCGGAGG 2279856  
 QY 824 oSerProSer----- 827  
 Db 2279857 GTACCCAGCTGTGGACGCGCTGTACGCGGCGCGCGGTTCGCGGGTGTACCGCGGAGAC 2279916  
 QY 828 -----ProSerPro----- 830  
 Db 2279917 CTGGCTGGAACGGGAGGTGGCGCGCGGTCTCGGTGAGCTTCGGGTTCACGGGAGC 2279976

QY 831 -----SerProSerProSerArgSerProSerProSerAla-----Se 843  
 Db 2279977 TGGTCGGGATCCAATGGCAGACCGACCGCTTCAAACTCGGCGGTGAGAACTCGACCGGT 2280036  
 QY 843 rPro-----SerProSerSerProSerProSerProSerProSerProSerProSe 861  
 Db 2280037 GCCTGATCGAAGCGCGCA-CGCTCGCTCATCCGAGGCGCGCGCGCGCGCGCG 2280095  
 QY 861 rProThrProSerSerProValSerGlyGlyVallyValGlnTyrLysAenAsnAs 881  
 Db 2280096 CCCTCCGCTCTCCACCGCGGTGAGACCGTGCCTG-----GGCAA 2280137  
 QY 881 pSer-AlaProGlyAspAenGlnIleLysProGlyLeuGlnValValAenThrGlySerS 901  
 Db 2280138 CTCCCGCTCTGTGTCAGGCTTCGCGCAGTCCGGCTCGGCTCGGCGCGGCGAGTA 2280197  
 QY 901 erSerValasp 904  
 Db 2280198 CCGCGCACGAC 2280208  
 RESULT 4  
 US-10-156-761-1845  
 ; Sequence 1845, Application US/10156761  
 ; Publication No. US20030119018A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OMURA, SATOSHI  
 ; APPLICANT: IKEDA, HARUO  
 ; APPLICANT: ISHIKAWA, JUN  
 ; APPLICANT: HORIKAWA, HIROSHI  
 ; APPLICANT: SHIBA, TADAYOSHI  
 ; APPLICANT: SAKAKI, YOSHIYUKI  
 ; APPLICANT: HATTORI, MASAHIRA  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-262  
 ; CURRENT APPLICATION NUMBER: US/10/156,761  
 ; CURRENT FILING DATE: 2002-05-29  
 ; PRIOR APPLICATION NUMBER: JP 2001-204089  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; PRIOR FILING DATE: 2001-08-02  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 1845  
 ; LENGTH: 2646  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces avermitilis  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)..(2646)  
 US-10-156-761-1845  
 Alignment Scores:  
 Pred. No.: 2,25e-164 Length: 2646  
 Score: 2478.00 Matches: 473  
 Percent Similarity: 64.07% Conservative: 128  
 Best Local Similarity: 50.43% Mismatches: 227  
 Query Match: 48.26% Indels: 110  
 DB: 15 Gaps: 14  
 US-09-917-376-1 (1-957) x US-10-156-761-1845 (1-2646)  
 QY 23 ThrAlaSerPheAlaValAlaAlaLeu-----GlyValLeuProIleAlaIleThr 40  
 Db 22 ACGGCGGTCTCGCTGGCGCGGTCTGCCCGCGGACCCCGCGGACCTGGCGCGG 81  
 QY 41 AlaSerProAlaHisAlaAlaThrThrGlnProTyrThrTrpSerAenValAlaIleGly 60  
 Db 82 AGCGCGCCCGCGCGGACGATCGCGCGGACAGTACAGCTGGAAGAACGCCCGGTGAC 141  
 QY 61 GlyGlyGlyPheValaspGlyIleValPheAenGluGlyAlaProGlyIleLeuTyrVal 80  
 Db 142 GCGCGCGGTCTCGTCCCGGATCGTCTTCAACCGCTCCGAGAGAACCTCGCTACGCC 201  
 QY 81 ArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaAenGlyArgTrpIleProLeu 100



202	CGCACCGACATCGCGCGCGCTACCGCTGGCGCGAGTCTCTCGAAGACCTTGGACCGCGCTG	261
101	LeuAspTrpValGlyTrpAsnAenTnpGlyTrpAsnGlyValValSerIleAlaAAsp	120
262	CTGACTCGGTTCGCTGGAGGAGTCTGGGGGACACCGGGTGTCTGTAGACCTTCGGCTCGAC	321
121	ProIleAsnThrAsnLysValTrpAlaAlaValIcIyMetTyrThrAsnSerTrpAspPro	140
322	TCCGTTCGACCCGAACAAGGTGTACGGCGCGCTCGGCACGTACACGAACAGCTGGGACCCG	381
141	AsnAspGlyAlaIleLeuArgSerAspGlnGlyAlaThrTrpGlnIleThrProLeu	160
382	GGCAACCGTCCGTCTAGTCCGGCGACCGGGGGCGAGCTGGCAGAAAGACCGACCTG	441
161	ProPheIysLeuGlyGlyAsnMetProGlyVArgGlyMetGlyGluArgLeuAlaValAsp	180
442	CCCTTCAAGCTGGCGGGGAAACATGCCGGGCGGGGACATGGCGGACCGGCTCGCGGTGCAC	501
181	ProAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSer	200
502	CCGACAGGAACAGCTGTGTATCTCGGCGCGCCACGCGCAAGGGCTGTGGCGGTGC	561
201	ThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIle	220
562	ACGACTCGGGGCGCTCTCGTCCGAGGTCAACGACTTCCCGAACGTCGCGACCTACGTG	621
221	AlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAla	240
622	CAGACCGGACCCACACAGCGGGGTACGCTCCGACAAACAGGCGATCTGTGTGGGTCA	681
241	PheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAla	260
682	TTTGACAGTTCAGCGGGTTCGCGGGAGCTCCACGGCGGAGCTGTACGTTCGGGTGC	741
261	AspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValPro	280
742	GACAAAGAACATCCGTCTATCGCTCCACGACGCGGGCGCGACTGTCTCCGGCTGCC	801
281	GlyAlaProThrGlyPheIleProHisIysGlyValPheAspProValAsnHisValLeu	300
802	GGCCAGCCCAACCGGCATCTCGGCCACAAAGGGCGTGTGGACGCGCGCAACGGCTGT	861
301	TyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLys	320
862	TACTTCGTGTACAGCGACAGGGCGGACCTGTACGACGGCGCGCAAGGACAGCTGTG	921
321	PheSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAspThrAla	340
922	TACACGACGAAGACCGGACCTGGACGAACATCAGCCCGGTTCGGGAGGCGCACAC	978
341	AsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMet	360
979	-----TACTACGGCTTCAGCGGGCTGACCGTGGACCGGACGACATCCGGGACCGGTG	1032
361	ValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGly	380
1033	CGGACTCGGTACAGCTCTGTGTGGCCCGGACACGACAGCTCTTCCGCTCCACGGACAGCGGC	1092
381	AlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrVal	400
1093	GGCACCTGGACGAAGGCGTGGGACTACACCTCGTATCCGACCGCTCGAACCGCTTCA	1152
401	LeuAspIleSerAlaGluProTrpTrpLeuThrPheGlyValGlnProAsnProProValPro	420
1153	ATGGATGTCGTCTCTCGCCCTTGCTCACTGGGGAGGCAACCCCGCACCGCCCGACGAG	1212
421	SerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg	440
1213	ACCCGGAACCTCGGCTGGATGACCGAGTCCCTGGAGATCGACCTGTCGACTCCCGCGGC	1272
441	MetLeuTyrGlyThrGlyValAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSer	460







```
Db 1219 GACCCACCGACGACCACTGGCTCTACGCGACCGGAATGACAATCTTTGGCGCCAC 1278
QY 454 AspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValIleGlyLeu 473
Db 1279 GATCTCACCACACTGGGACGACGCGCACCAATGTGTAATCCAATCAATCATGCGACGCGCATC 1338
QY 474 GluGluThrAlaValAsnAspLeuIleSerProSerGlyValAlaProLeuIleSerAla 493
Db 1339 GAGGAATCTCTCGTCAGGACCTGGCTCTGACCGCGGGAAGCAGCTATTGGCCGCA 1398
QY 494 LeuGlyAspLeuGlyGlyPheThrHisAla-----AspValThrAlaValProSer 510
Db 1399 GTGAGAGACGACCAACGGCTTACCTTTCAGCAGCAAGAACGACTCGGACATCGCGCAG 1458
QY 511 ThrIlePheThrSerProValPheThrThrGlyThrSerValAspTrpAlaGluLeuAsn 530
Db 1459 ACGGTCTGGGCAACGCGCCACATGGCCACCTCGACGAGGTGCACTACGCGCGGAACGTCG 1518
QY 531 ProSerIleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHis 550
Db 1519 GTCAAGAGCGTCTCGCTCGCTCGCAACACCGCGCGCACCAACAG-----1563
QY 551 ValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyVal 570
Db 1564 GTGGCCATCTCTCGACGCGCGCGGAGCTGAGCATCGACTACGCGCGCGACAGCTCC 1623
QY 571 ThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGly 590
Db 1624 ATCAACGGCGGCGACGCTGCTTATTCGGCGGCGGCGACAGCATCTCTGTGTCGACCGCC 1683
QY 591 AspProGlyGlnProValValTyrrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGln 610
Db 1684 TGTCTCGGC-----GTGACGCGCTCGCATCTCCAGGCGAGCTTGTCTCTCGCTCG 1734
QY 611 GlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrrAla 630
Db 1735 AGCTGCGCGCGCGCGCGCTCATCGCTCGGACCAAGAACACCAAGCGTCTTCTACGCG 1794
QY 631 LeuSerAsnGlyThrPheTyrrArgSerThrAspGlyGlyValThrPheGlnProValAla 650
Db 1795 GGTCTCGGATCGACCTTTTACGTACGACGAGGACACCGGCGACAGCTTC-----ACG 1845
QY 651 AlaGlyLeuProSerSerGlyAlaValGlyValMet-----PheHisAlaValProGly 668
Db 1846 CGCGG-----CCCAAGCTGGCGACGCGGACGATCGCGGATATCGTCTCACCGGAC 1902
QY 669 LysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrrHisSerThrAsnGlyGly 688
Db 1903 ACGCGGGGACGCTGTATGTCTCGACCGACGTCGCGCATATTCGCTCCACAGACTCGGGC 1962
QY 689 SerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsnValGlyPheGlyLys 707
Db 1963 ACGACCTTTGGCAAGTCTCCACCGCTCGACCAACACCTACGATCGCGCTGGGTGTCG 2022
QY 708 SerAlaProGlySerSerTrpProAlaValPheValGlyThrIleGlyValThr 727
Db 2023 GGCTCA---GGCTCGAAGTGG---AACCTGTATGCTCTCGGCACC-----GGCCGCTCA 2070
QY 728 GlyAla-----TyrrArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAspApp 745
Db 2071 GGGGCTCGCTCTACGCCAGTGGAGACAGCGGCGCTCTCTGGACGGACATCCAGGGCTCC 2130
QY 746 GlnHisGluTyrrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArg 765
Db 2131 CAGGGTTCGGTCCATCGACGACCAAGGTGCGCGGACGCGGACCGCGCGGCGCA 2190
QY 766 ValTyrrIleGlyThrAsnGlyArgGlyIleValTyrr-----GlyAspIleGlyValAla 783
Db 2191 GTCTAGTGGGACCAACGCGCGGGCGTCTTTTACGCTCAGGGAACCGTCTCGGCGGCG 2250
QY 784 ProSerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerProSer 803
```

```
Db 2251 ACGGGCGGAGCT-----TCCTCGTCGACCAACGACGAGC 2283
QY 804 ProSerProSerSerSerProSerProSerProSerProSerProSerProSerSer 823
Db 2284 AGCAGCAGTACCTCTTCCGCCAGCTCGAGCACCAACGCTGAGGTGAGGTTGTATCCACG 2343
QY 824 ProSerProSerProSerProSerProSerProSerProSerProSerProSerProSerAlaSe 843
Db 2344 ACCCGGGCTTCGACGGTGACTTCGTCAGGACCACTCGGCGCGCTCCACGCGGTCA 2403
QY 843 rProSerPro 846
Db 2404 GGGGTGCGCG 2413

RESULT 6
US-10-026-994-1
; Sequence 1, Application US/10026994
; Publication No. US20030113732A1
; GENERAL INFORMATION:
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Goedegebuur, Frits
; APPLICANT: Ward, Michael
; APPLICANT: Yao, Jian
; TITLE OF INVENTION: EGV1 Endoglucanase and Nucleic Acids
; FILE REFERENCE: GC698
; CURRENT APPLICATION NUMBER: US/10/026,994
; CURRENT FILING DATE: 2002-04-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2710
; TYPE: DNA
; ORGANISM: Trichoderma reesei
US-10-026-994-1

Alignment Scores:
Pred. No.: 6,15e-105 Length: 2710
Score: 1631.50 Matches: 353
Percent Similarity: 57.23% Conservative: 130
Best Local Similarity: 41.82% Mismatches: 297
Query Match: 31.77% Indels: 64
DB: 15 Gaps: 23

US-09-917-376-1 (1-957) x US-10-026-994-1 (1-2710)
QY 25 SerPheAlaValAlaAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAla 44
Db 50 TCTCGAGTCTTGGCTTGTCTTGGGCGCGTC-----ATCCCTGCC 91
QY 45 HisAlaAlaThrThrGlnProTyrrThrTrpSerAsnValAlaIle---GlyGlyGlyGly 63
Db 92 CATGCTGCC-----TTTTCATGGAAGAACGTCAAGCTCGGCGCGCGCGCGC 139
QY 64 PheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrrValArgThrAsp 83
Db 140 TTGTCGCCCCGGCATCATCTTCCATCCCAAGAAAAGCGTAGCATATGCACCAACAGAT 199
QY 84 IleGlyGlyMetTyrrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrp 103
Db 200 ATTGGCGGGCTGTACCGCTCAAC---GCCGACGACTCATGACCGCGCTCAGGATGG 256
QY 104 Val-----GlyTrpAsnAsnTrpGlyTyrrAsnGlyValValSerIleAlaAla 119
Db 257 ATTGCTGATAATGCGGCTGCGCAACTGG-----GGCATCGACGCTGTGGCTT 307
QY 120 AspProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrrThrAsnSerTrpAsp 139
Db 308 GATCCGAGGACGATCAAAAGGTGTATGCGCAGTCGCGCATGTATACGAACAGCTGGAT 367
QY 140 ProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrPro 159
Db 368 CCGAGTAATGAGGCCATCATCTGCTCGTCAGACCGCGGCGCAACGCTGTCTTCCACCAAC 427
```







US-10-156-761-1/c  
 ; Sequence 1, Application US/10156761  
 ; Publication No. US20030119018A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OMURA, SATOSHI  
 ; APPLICANT: IKEDA, HARUO  
 ; APPLICANT: ISHIKAWA, JUN  
 ; APPLICANT: HORIKAWA, HIROSHI  
 ; APPLICANT: SHIBA, TADAYOSHI  
 ; APPLICANT: SAKAKI, YOSHIYUKI  
 ; APPLICANT: HATTORI, MASAHIRA  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-262  
 ; CURRENT APPLICATION NUMBER: US/10/156,761  
 ; CURRENT FILING DATE: 2002-05-29  
 ; PRIOR APPLICATION NUMBER: JP 2001-204089  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; PRIOR FILING DATE: 2001-08-02  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 1  
 ; LENGTH: 9025608  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces avermitilis  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (4187715)  
 ; OTHER INFORMATION: a, t, c, g, other or unknown  
 US-10-156-761-1

Alignment Scores:  
 Pred. No.: 6,1e-97 Length: 9025608  
 Score: 1570.00 Matches: 393  
 Percent Similarity: 49.25% Conservative: 131  
 Best Local Similarity: 36.94% Mismatches: 346  
 Query Match: 30.57% Indels: 194  
 DB: 15 Gaps: 38

US-09-917-376-1 (1-957) x US-10-156-761-1 (1-9025608)

QY 4 SerGluAAsnIleArgLeuThrMetArg-----SerArgArgLeuValSerLeu 19  
 DB 3162026 TCTCCAGACGAGGTATGTCTATCGCAGCCGCCGCCGAGCAGACCGTCTC----- 3161973  
 QY 20 LeuAlaAlaThrAlaSerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIle 39  
 DB 3161972 CTCGCGGGACCGCGCG-----GCCGCGCGCTCACGGCGTCTCC-----GCC 3161928  
 QY 40 ThrAlaSerProAlaHisAlaAlaThrThr-----GlnProTyrThrTrpSerAsn 56  
 DB 3161927 GTGGGGTCTCCAGCGCAGCCGCCGAAACACCGCGCGCGCTCTGTACCGCTGGCGCAAC 3161868  
 QY 57 ValAlaIleGlyGlyGlyPheValAlaAspGlyIleValPheAsnGluGlyAlaProGly 76  
 DB 3161867 GCCGTATCGGGGACCGCGCTCTGTACCGCGGTCTCTCCACCCCTCCGTACCGGT 3161808  
 QY 77 IleLeuTyrValArgThrAspIleGlyMetTyrArgTrpAspAlaAlaAsnGlyArg 96  
 DB 3161807 CTCGCTACCGCGGACCGACATCGCGCGCTACCGCTGGGACGACCGCGCGCGCCG 3161748  
 QY 97 TrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValSer 116  
 DB 3161747 TGGACCCCGCTCATCGACCACTCGCGCTGGGACGACTGGAACCTCTCTCGCGCTCGAGGCG 3161688  
 QY 117 IleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsn 136  
 DB 3161687 ATGGCGGTGACCCACGACCGCGGCTCTACTCGCGCTGGGACCTACGCGCCAG 3161628  
 QY 137 SerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGln 156  
 DB 3161627 TGTGGCGGGACAC-----GGCGGGTCTCTGCGCTCGAGGACCGCGGCGCCACCTGGACC 3161571  
 QY 157 IleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArg 176

DB 3161570 CGCACCGACCTGACCGTGAAGCTCGCGCGCAACGAGACGCGCGCGCGGTGAGCGA 3161511  
 QY 177 LeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGly 196  
 DB 3161510 CTCCTCGTCGACCGCGCGCAGCAGCAGCACCTCTGCTGGGACG---CGGACGACGGG 3161454  
 QY 197 LeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspVal 216  
 DB 3161453 CTGCTCAAGTCGACCGCGCGCGCGCTTGGGACGCGCGACCGCTTCCCG----- 3161400  
 QY 217 GlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyVal 236  
 DB 3161399 -----GCGAAGCGGAACCTCTCTCGGCGGAGGAGTC 3161370  
 QY 237 ValTrpValAlaPheAspLysSerSerSerSerLeuGlyGlnAlaSerLysThrIlePhe 256  
 DB 3161369 GTGTTT-----CTCGTCGCCCGCGCGCGCACCGTCTAC 3161337  
 QY 257 ValGlyValAlaAspProAsnAsnPro-----ValPheTrpSerArgAsp 271  
 DB 3161336 GCGGCTGGGTGACGCGCAGCGCACCTCGGGCAGCGGACCTGTACCGCAGCGCGCAC 3161277  
 QY 272 GlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe-----IlePro 288  
 DB 3161276 ---GGCAGCACCTGGGGGCGCTCCCGCGCGCGCTCGGACCTCGGCAAGTCCCG 3161220  
 QY 289 HisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGly 308  
 DB 3161219 CTCGCGCGCGGTACGACACGACCGCGCGCTGTACGTACGTCAGCGGACGCGCACCC 3161160  
 QY 309 GlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGln 328  
 DB 3161159 GCGCGCGGGCGCGCAGTCCGACGCGCAGGTGCACAGCTGCGTACCGCCACGCGGACGTGG 3161100  
 QY 329 ThrArgIleSerProVal-----ProSerThrAspThrAlaAsnAspTyrPhe 344  
 DB 3161099 ACCGAGGTACCCCGGTGAAGCGCGGCGGACGACGACGACGCGTCCGGGACACCTTC 3161040  
 QY 345 GlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGln 364  
 DB 3161039 GCCTACGCGGGGTCTCGCGTCCGACCGCGCGCGCGCGCGCTCTCCACCAAC 3160980  
 QY 365 IleSerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThr 384  
 DB 3160979 AACCGCTGGGCGCAGCGGACACGGTCTTCGGGTCCACGACGCGCGCGCTACTGACG 3160920  
 QY 385 ArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSer 404  
 DB 3160919 TCCTCAAGAC-----GCCGCGGTTCGACGCTGTC 3160887  
 QY 405 AlaGluProTrpLeuThrPheGlyValGlnProAsnProValProSerProLysLeu 424  
 DB 3160886 GAGACTCCCTTCCTCGACTGGGCGCGAC-----AAGCGGAAGTTC 3160845  
 QY 425 GlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGly 444  
 DB 3160844 GCGTGTGGATCCAGCGCTCGCGGTCCGCGTACCGCTCCAGCAGCGCTGTGTACCGG 3160785  
 QY 445 ThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlnIle 464  
 DB 3160784 ACCGGCGGACCTCTACGCGCACCGCGACTCAAGGCTGG----- 3160743  
 QY 465 HisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerPro 484  
 DB 3160742 -----GCACCGCGGATCCCGCGCTCGAGGAGAGCGCGCTGCCCACTGATCTCGCCC 3160689  
 QY 485 ProSerGly---AlaProLeuIleSerAlaIleuGlyAspLeuGlyGlyPheThrHisAla 503  
 DB 3160688 CCGGTCCGGGAGGACACCTGATCAGCGGACTCGGGGACATCGGTGTGTATGATACCGAG 3160629  
 QY 504 AspValThrAlaValProSerThr---IlePheThrSerProValPheThrThrGlyThr 522



Db 3160628 CGGCTCACGGCGTCTCGTTCGGCGGATGCGGACGAAACCCCGTTCGGGTGCGCGACG 3160569

Qy 523 SerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspPro 542

Db 3160568 GGACTCGCGCAGCGCGCGCCAGCGCGCGTATGTCGCGACACGGGTGGGGCGACAC 3160509

Qy 543 SerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTyr--- 561

Db 3160508 GGCACAGCGC-----GCGTACTCCACAGCAGCGGGCGGCGACCTTGGCG 3160467

Qy 562 ---PheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAla 580

Db 3160466 CCCTTCGAGGCCAGCGCGACATGCCAAGACGACCGGGACCGATCGCCACCAAGTGCC 3160407

Qy 581 AspGlySerArgPheValTyrAla-----ProGlyAspProGlyGlnProVal 596

Db 3160406 GACGGCGGCACACATGCTGTGTGCTCTCGTGCACCTGGGACGCGC-----ACG 3160362

Qy 597 ValTyrAla-----ValGlyPheGlyAsnSerTyrAlaAlaSerGlnGlyVal 612

Db 3160361 ACGTACGGCGGCCACCGCTCGACGACACGCGCGAGTGTGTCGAGGTCTCTCTCTTC 3160302

Qy 613 ProAlaAsnAlaGlnIleArgSerAspArgValAsnProTyrThrPheTyrAla----- 630

Db 3160301 CCGAAGGCGCGCACACCGGTGCGCGACCGCGCGCATCGACGCGCTTACGCTCAGCAG 3160242

Qy 631 LeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAla 650

Db 3160241 TTCGACAATGGAACGCTATACGCCAGCAGCTGACAGTGGCGGTTCGTCACGCGCGGTGCG 3160182

Qy 651 AlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysGlu 670

Db 3160181 GCGGACTGCCCC---TCCGGCAGACGACGATTCAGGTGTCGCGCGCGCGGACGAGC 3160125

Qy 671 GlyAspLeuTyrLeuAlaAlaSer---SerGlyLeuTyrHisSerThrAsnGlyGlySer 689

Db 3160124 GCGGACCTGTGGCTCTCCGCCAAATGGAACGGGTCTACCGGTCCACGACGCGGGGAC 3160065

Qy 690 SerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAla 709

Db 3160064 ACCTTCGCGAGATCGACAGCTGTGCGGCTCGTACACCTCTCGGCTTCGGCAAGCGGCG 3160005

Qy 710 ProGlySerSerTyrProAlaValPheValGlyThrIleGlyGlyValThrGlyAla 729

Db 3160004 GACGCGCGGACTACCGCGCATCTACAGGTGCGTCCGACGAGACCATCACCGCGGTC 3159945

Qy 730 TyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAspAspGlnHisGlnTyr 749

Db 3159944 TACCGCTCCGACGACGCGCCGAGCATGGTTCGGATCAACGACGACGCGCCACCATGG 3159885

Qy 750 GlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgValTyrIleGly 769

Db 3159884 GGGTGGATCGCGGAGCGGTGCGGTGACCGCGCATCCACGCGCGGTCTACCTCGCC 3159825

Qy 770 ThrAsnGlyArgGlyIleValTyrGlyAspIle-----GlyGlyAlaPro 784

Db 3159824 ACCAAGCGCGCGCATCTACGATAGGGGCGCGTTCGATCCCGGGGTCTACGGATGCCA 3159765

Qy 785 SerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerProSerPro 804

Db 3159764 CCGGGCGCGCATCTGTTCGGTGGC---ACTACACCCCGACGAGTGGCGCGGAGG 3159708

Qy 805 Ser-----ProSerSerProSerPro----- 812

Db 3159707 TCTGCGAGGAGCGATACGCTCTCATGAGGAGCGCGGTCACTCCGTACCGTCCGCGC 3159648

Qy 813 ---SerProSerProSerSerPro----- 820

Db 3159647 TCTTCTCTGGGCGGAGCTCGAACCCACCGCGGGCGGAGGAGTTCGGGTGCGTCCGAC 3159588

Qy 821 -----SerSerSerProSerPro----- 828

Db 3159587 GGCTCATGGACCTGATGACAGAGAACGGCATCGGGGTGCTCTGGCCACCCCGCCGCT 3159528

Qy 829 SerProSerPro----- 832

Db 3159527 CGCCCCCGCTTGGATGGGGCGCTGACACCCGAGACCTGCCCCCGACGAGGACGGCC 3159468

Qy 833 -----SerProSerArgSerProSerProSerAla 842

Db 3159467 GCATCGAGTGGTGGGGCGCGCCAGCAGCTTCGCCCACTCCAGTGCCTTACCGCGCT 3159408

Qy 843 SerProSerProSerSerProSerProSerProSerSer-----ProSerSer 858

Db 3159407 ACGCCGCGCGCATCACCGAGGACCTTCGCCCGCGCTACGCGGGCCATCCCGCCCTCACGA 3159348

Qy 859 SerProSerProThrProSerSerProValSerGlyGlyValLys----- 874

Db 3159347 TGTGGCAGATCAACACAGAGTACTGACCTAGCAGTGGGGCGGACGAGGGCGCCACCGCCT 3159288

Qy 875 -----ValGlnTyrLysAsnAspSerAlaProGlyAsp 886

Db 3159287 TCCGCGCGTGGCTCCAGGGGAGGTACGGCAGCGCTCGACGCACTCAACACGCGCTGG--- 3159231

Qy 887 AsnGlnIleLysPro---GlyLeuGlnValValAsnThrGlySerSerSerValAspLeu 905

Db 3159230 -----GCACGCGCTTCTGAGCCAGGCTTACGACACTGGGAAGGCATCTCCCGCCGC 3159177

Qy 906 SerThrValThrValArgTyrTyrPheThrArgAspGlyGlySerSerThrLeuValTyr 925

Db 3159176 GCCAGCCCACTACATGAAACACCCACCCAGGTGCTCGACTTCAAGCGCTTACCTCCG 3159117

Qy 926 AsnCysAspTyrAlaAlaIleGlyCysGlyAsnIleArgAlaSerPheGlySerValAsn 945

Db 3159116 ACGCACTCATGGAGTGCT---TCGTCGCGGAAGGTGACATCGTCG----- 3159075

Qy 946 ProAlaThrPro 949

Db 3159074 CCGCGCACACCC 3159063

RESULT 8

US-10-156-761-2561

; Sequence 2561, Application US/10156761

; Publication No. US20030119018A1

; GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI

; APPLICANT: IKEDA, HARUO

; APPLICANT: ISHIKAWA, JUN

; APPLICANT: HORIKAWA, HIROSHI

; APPLICANT: SHIBA, TADAYOSHI

; APPLICANT: SAKAKI, YOSHIYUKI

; APPLICANT: HATTORI, MASAHIRA

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCE: 249-262

; CURRENT APPLICATION NUMBER: US/10/156,761

; CURRENT FILING DATE: 2002-05-29

; PRIOR APPLICATION NUMBER: JP 2001-204089

; PRIOR FILING DATE: 2001-05-30

; PRIOR APPLICATION NUMBER: JP 2001-272697

; PRIOR FILING DATE: 2001-08-02

; NUMBER OF SEQ ID NOS: 15109

; SEQ ID NO 2561

; LENGTH: 2217

; TYPE: DNA

; ORGANISM: Streptomyces avermitilis

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(2217)

US-10-156-761-2561

Alignment Scores:

Pred. No.: 1e-91 Length: 2217

Score: 1442.00 Matches: 328

Percent Similarity: 54.17% Conservative: 107

Best Local Similarity: 40.85% Mismatches: 272

Query Match: 28.08% Indels: 96



DB: 15 Gaps: 25  
US-09-917-376-1 (1-957) x US-10-156-761-2561 (1-2217)  
QY 7 IleArgLeuThrMetArgSerArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPhe 26  
    |||||  
DB 1 ATGGCGACGCCGCCGCCGAGCAGACCGCTC-----CTCGCGGAGACCGCGCG-- 51  
QY 27 AlaValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProIleHisAla 46  
    |||||  
DB 52 -----GCCGCGCGCTCACGCCCTCC-----GCCGTGGGTGCCAGCGCACGCC 99  
QY 47 AlaThrThr-----GlnProTyrThrTTPSerAsnValAlaIleGlyGly 63  
    |||||  
DB 100 GCGGAACACCGCGCGCGCTGCTACCGCTGGCGCAACCGCGTCTATCGGGGACCGGC 159  
QY 64 PheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAsp 83  
    |||||  
DB 160 TTCGTACCGGGGTCTCTTCACCCCTCCGTACCGGTCTCGCTACGCCCGGACCGAC 219  
QY 84 IleGlyMetTyrArgTTPAspAlaAlaAsnGlyArgTTPileProLeuLeuAspTTP 103  
DB 220 ATCGCGCGCGCTACCGCTGGACACCGCGCGCGCTGACCGCTCATCGACACAC 279  
QY 104 ValGlyTTPAsnAsnTTPGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsn 123  
DB 280 CTCGGCTGGGACGACTGGAACCTCTCGCGCTCGAGGCGATGGCGTCCAGCCCGAC 339  
QY 124 ThrAsnLysValTTPAlaAlaValGlyMetTyrThrAsnSerTTPAspProAsnAspGly 143  
DB 340 CCGGACCGCGTCTACTCGCGCTGGACCTACGCGCCAGTCTGGCGGGGCAAC--GGC 396  
QY 144 AlaIleLeuArgSerSerAspGlnGlyAlaThrTTPGlnIleThrProLeuProPheLys 163  
DB 397 GCGGTCTCGCTCCGAGACCGCGCGCACCTGGACCCCGACCGACCTGACCGGTGAAG 456  
QY 164 LeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsn 183  
DB 457 CTCGGCGGCAACAGGACCGCGCGCGCGCTGGAGCGACTCTCTGTCACCGCGCGAC 516  
QY 184 AspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSerThrAspSer 203  
DB 517 AGCGACCCCTCTGCTGGGACG---CGGACGACGGGCTCTCAAGTCGACCGACCGG 573  
QY 204 GlyAlaThrTTPSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnPro 223  
DB 574 GCGGCCACTTGGGACCGCGACCGCTTCCG-----  
QY 224 ThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTTPValAlaPheAspLys 243  
DB 607 -----GCCAAGGCGCAACTCTCTCGGCGCAGGAGTCTGTTC----- 642  
QY 244 SerSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsn 263  
DB 643 -----CTGTCGCCCGCGCGCGACCGTCTACCGCGCTGGGGTACGCGCGAC 690  
QY 264 AsnPro-----ValPheTTPSerArgAspGlyGlyAlaThrTTPGlnAla 278  
    |||||  
DB 691 GGCACCTCGGCGACCGGCAACCTGTACCGCACCGCGAC---GGCAGCAGCTTGGGGGCC 747  
QY 279 ValProGlyAlaProThrGlyPhe-----IleProHisLysGlyValPheAspPro 295  
DB 748 GTCCCCGCGCGCGCTCCCGCACCTCCGCGCAAGGTCTCCGCTCGCGCGCGTACGACAG 807  
QY 296 ValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSer 315  
DB 808 CACACCGCGAGCTGTACGTACGTACGCGACGACCGCGCGCGCGCGCGCGCGCGCG 867  
QY 316 GlyAspValTTPLysPheSerValThrSerGlyThrTTPThrArgIleSerProVal--- 334  
DB 868 GCGACGTGTACAAAGCTGGTACCGCACCGGACCGTGGACCGAGGTACCCCGGTGAAG 927  
QY 335 -----ProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIle 351

DB 928 CCGCGCGGACGACGACGACGCGCTCGGCGGACACCTTTCGCTACGCGGGGTGCGCGTC 987  
QY 352 AspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTTPTrpProAspThr 371  
DB 988 GAGCGCGCGCGCGCGCGCACCTCGTGTCTCCACCAACAACCGCTGGCGGACGCGAC 1047  
QY 372 IleIlePheArgSerThrAspGlyGlyAlaThrTTPThrArgIleTTPAspTTPTrpSer 391  
DB 1048 ACGGTCTTCGGTCCACGACGCGCGCGCTGACCTGGACGTCCTCAAGGAC----- 1098  
QY 392 TyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTTPLeuThrPhe 411  
DB 1099 -----GCCGCGGTGTTCGACGTGTCCGAGACTCCCTCTCTCGACTCGG 1140  
QY 412 GlyValGlnProAsnProValProSerProLysLeuGlyTTPMetAspGluAlaMet 431  
DB 1141 GCGGACGAC-----AAGCCGAAGTTCGGCTGGTGGATCCAGCGCTC 1182  
QY 432 AlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAla 451  
DB 1183 GCGGTCCGACCGGTACGACTCCCGACGCTGTATCGGACCGCGCGGACCTCTACGCG 1242  
QY 452 ThrAsnAspLeuThrLysTTPAspSerGlyGlyGlnIleHisIleAlaProMetValLys 471  
DB 1243 ACCCGGACCTCAAGCGCTG-----GCACCGCGGATCGC 1278  
QY 472 GlyLeuGluGluThrAlaValAsnAspLeuIleSerProProSerGly---AlaProLeu 490  
DB 1279 GCGCTCGAGAGAGCGCGCTGCGCAACTGATCTCGCCCGCGTTCGGGAGGACACCTG 1338  
QY 491 IleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSer 510  
DB 1339 ATCAGCGGACTCGGGGACATCGGTGTGATGATCCAGGCGGCTCACGGGTCTCCGTG 1398  
QY 511 Thr---IlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeu 529  
DB 1399 CCGCGCATGGCAGCAACCCCGTTCGGGTCCGCGACGGGACTCGCGACGCGCGCGCC 1458  
QY 530 AsnProSerIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArg 549  
DB 1459 AGCGCGCGTATGTCGTCGCGACGGGTGGGGCGACCAACCGCAACGCG-- 1506  
QY 550 HisValAlaPheSerThrAspGlyGlyLysAsnTTP-----PheGlnGlySerGluPro 567  
DB 1507 -----CGGTACTCCACGACCGCGCGCGACTTGGCGCGCTTCGAGCGCCAGCCGAC 1560  
QY 568 GlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTTP 587  
DB 1561 ATCGCCAGGACGACCGGACCGATGCCACCGACCGGCTCCGACGCGGCGGACACTGCTG 1620  
QY 588 Ala-----ProGlyAspProGlyGlnProValValTyrAla----- 599  
DB 1621 TCCTTCGTCACCTGGGACGCG-----ACGACGTACGCGCGCCACCGCTCG 1665  
QY 600 ValGlyPheGlyAsnSerTTPAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArg 619  
DB 1666 ACGGACAACCGCGCGAGCTGTCTCCCTTCCTCCGAAAGGCGCGCACCGGTC 1725  
QY 620 SerAspArgValAsnProLysThrPheTyrAla-----LeuSerAsnGlyThrPheTyr 637  
DB 1726 GCGACCGCGCGCTTCGACCGCGCTTACGCTACGACTCGACATCGACATGGACGCTATAC 1785  
QY 638 ArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGly 657  
DB 1786 GCCAGCAGTACAGTGGCGGTTCGTTCCAGCGCGCTGCGGCGGAGCTGCCCC--TCCGCG 1842  
QY 658 AlaValGlyValMetPheHisAlaValProGlyLysGluGlyAspLeuTTPLeuAla 677  
DB 1843 GACAGCGAGTTCAGCTGTGCGCGCGCGCGGACGAAGCGGACCTGTGTCTCTCCGCG 1902  
QY 678 Ser---SerGlyLeuTyrHisSerThrAsnGlyGlySerSerTTPSerAlaIleThrGly 696



Db 1903 AATGGAACGGCTCTACCGGTCCACGACGGCGGGACACCTTCGCGAGGATCGACAGC 1962  
Qy 697 ValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAla 716  
Db 1963 TGCTGGGCTCGTACACCTCGCTTCCGGAAGCGCGCGACGGCGCGACCTATCCCGGCG 2022  
Qy 717 ValPheValValGlyThrIleGlyGlyValThrGlyAlaTyrArgSerAspAspCysGly 736  
Db 2023 ATCTACCGGTGGCTCGACGGAGACCATACCGCGCTCTACCGCTCGACGACGGCGCC 2082  
Qy 737 ThrThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIle 756  
Db 2083 AGGACATGGGTCCGGATCAACACGACGCCACACAGTGGGGTGGATCGCGAGCGCGTC 2142  
Qy 757 ThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleVal 776  
Db 2143 GTCGGTACCGCGCATCCACGGCGGGTCTACCTCGCCACCAACGGCGCGGCATCCAG 2202  
Qy 777 TyrGlyAsp 779  
Db 2203 TACGGGAG 2211

## RESULT 9

US-09-927-827-21  
; Sequence 21, Application US/09927827  
; Publication No. US20030036176A1  
; GENERAL INFORMATION:  
; APPLICANT: Bower, Stanley G.  
; APPLICANT: Ramseier, Thomas M.  
; TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris  
; FILE REFERENCE: 38-10(15824)B  
; CURRENT APPLICATION NUMBER: US/09/927,827  
; PRIORITY FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: US 60/279,493  
; PRIORITY FILING DATE: 2001-03-28  
; NUMBER OF SEQ ID NOS: 69  
; SEQ ID NO 21  
; LENGTH: 3668  
; TYPE: DNA  
; ORGANISM: Xanthomonas campestris  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1001)..(2668)  
US-09-927-827-21

Alignment Scores:  
Pred. No.: 3,16e-72 Length: 3668  
Score: 1167.50 Matches: 299  
Percent Similarity: 48.55% Conservatives: 137  
Best Local Similarity: 33.30% Mismatches: 351  
Query Match: 22.74% Indels: 112  
DB: 10 Gaps: 28

US-09-917-376-1 (1-957) x US-09-927-827-21 (1-3668)

Qy 23 ThrAlaSerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSer 42  
Db 1136 ACCAACAGGAGCATGGCTGGTGTCTTGAGCTGCTG---CTGTGCTCTTTTCGACG 1192  
Qy 43 Pro-----AlaHisAlaAlaThrGlnProTyrThrTrpSerAsnValAla 58  
Db 1193 CCCAGCGTGGCGCGCGAGCCCGCCAGCTCCGGGCCCTACAGTGGCGCAGTGTGCGC 1252  
Qy 59 IleGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeu 78  
Db 1253 ATTGGCGGTGGCGGCTTTGTACCGGTGTGCTGTTTCATCCCGCGGACGCTGTGCGC 1312  
Qy 79 TyrValArgThrAspIleGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIle 98  
Db 1313 TATGCGCGCACCGATGGGTGGCGGTACCGCTGGATGCGAGCGCGCAGCAGTGGACC 1372  
Qy 99 ProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAla 118

Db 1373 CGCTGACCGAGCTGTTGGCGCTGACGACCTGGAACCTGATGGGCATCGACGATTCGCC 1432  
Qy 119 AlaAspProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrp 138  
Db 1433 GTGACACCGCGGATGCGGATGCGCTGTATCTGGGGCGCGGACCTATATGATGAGAA--- 1489  
Qy 139 AspProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThr 158  
Db 1490 CGCGCGGCGCATGCGCGCAGTGTGGCTCGTTCAACCGCGCGCGCACGTTTCGAGCGTGC 1549  
Qy 159 ProLeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAla 178  
Db 1550 GACCTGCGCGTTAAGCTGGTGGTAACACAGCTGGCGCGCGCAATGGCGAGCGGTGGCG 1609  
Qy 179 ValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrp 198  
Db 1610 GTGACCGCGCAGATGGCGCGGTGCTGTCTGTGGGCTCGCGCATGCC---GGCTGTGG 1666  
Qy 199 ArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThr 218  
Db 1667 CGTAGCGAGATCGCGCGCGCCTGGGCGAAGTGGCTGCTTCCGACGCGCGCTG 1726  
Qy 219 TyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrp 238  
Db 1727 GCGGTGCGCACCGCGCGCAATCATGTGGCGCGCAGCAGCGGTG---GGGATCGCTTT 1783  
Qy 239 ValAlaPheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGly 258  
Db 1784 GTCTGTTCGACGACGAGCAGTGGCAACAGCGCTGCCAACACCGCGCATCTACGTGGCG 1843  
Qy 259 ValAlaAspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAla 278  
Db 1844 GTGTCCCGCAACACGACCGCTGTATGTTCGAAGATGCCGCGCGAGTGGCGACCG 1903  
Qy 279 ValProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHis 298  
Db 1904 GTGGCGGGCAACACACGCGGCTGCGCGCGCACATGCGCGCGCGCAGCATGGCGCAC 1963  
Qy 299 ValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspVal 318  
Db 1964 ---TGTATCTGAGCTATGGCGACCGCGCGCGGAGCTGATGGCGGGGAGCGCTTG 2020  
Qy 319 TrpLysPheSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAsp 338  
Db 2021 TCGAAATTCACCGCGCGCACGGGCGCTGGCTGGATGAGATCAGCCCGATTCGCGAG---CCA 2077  
Qy 339 ThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThr 358  
Db 2078 GCCAGTGGCGATGGATTCGCTGGGTGGGTGGGTGGATCCGCAACATCCGCGAGGTG 2137  
Qy 359 IleMetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAsp 378  
Db 2138 CTGCTGGCGCAGCACCTTCGCGCTGCGACGCGCGCGCAGAGCTGTATCGACGCTGAT 2197  
Qy 379 GlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArg 398  
Db 2198 GTTGGCAAGCACTGACGCGCTG-----TTGGCGCAT 2230  
Qy 399 TyrValLeuAspIleSerAlaGluProTrpTrpLeuThrPheGlyValGlnProAsnProPro 418  
Db 2231 GCGGTGTTTCGATCACAGCGCGCGCGCTGG---ACCGCACATGCCACGCGCGCAC----- 2281  
Qy 419 ValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSer 438  
Db 2282 -----TGGATGGGG---GCGCTGGCGATCGATCCGTTTCGACGCGC 2317  
Qy 439 AspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrp 458  
Db 2318 AACCATGCGCTGTTCTGTGACCGGCTACCGCATCTGGGCTTCGCGCAATCTG----- 2368  
Qy 459 AspSerGlyGlyGlnIleHisIleAlaPro-----MetVal 470  
Db 2369 -----CAGGATTCGCGCGCACCGCGCGCTGCGTGGTGGTTCAGGAC 2416







Db 393 ATCGCGTGGACCCCAACACCCGACAGGCTGTACCTCGCGCAGCGGCGCTATGTCCGC 452  
Qy 136 AsnSerTrpAspProAsnAspGlyAlaIleuArgSerSerAspGlnGlyAlaThrTrp 155  
Db 453 GACGAGTGG-----GCGCGCTTCTATGTGTCGGAAGACCCGCGGCAGCTCGTTT 500  
Qy 156 GlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyValGlyMetGlyGlu 175  
Db 501 ACAATCTACGAGTCCGCTTCCGATGGCGCCACGACATGGAGCGCAACATGGCGAG 560  
Qy 176 ArgLeuAlaValAspProAsnAsnAspAsnIleuLeuTyrPheGlyAlaProSerGlyLys 195  
Db 561 CGCTCGCTGTCACACCGTTTCAACTCGAACGAGGCTTGGATGGGTACGCGTACA---GAG 617  
Qy 196 GlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAsp 215  
Db 618 GGTATCTGGAAGATTCGCGACCGCGCCCAAGACCTTGGACAAACGTCAACGTCATCCCGAC 677  
Qy 216 ValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGly 235  
Db 678 GCGTTC-----ACCAACGGTATCGGATACAGTGC----- 707  
Qy 236 ValValTrpValAlaPheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIle 255  
Db 708 -----GTCAATTTTCGACCCC-----GAACGTAAATGGCACCATC 740  
Qy 256 PheValGlyValAlaAspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThr 275  
Db 741 TACGCGACGCGACTCCCGCCAGCGC---ATGTACGTACGCGACGACGCGGTGTCTCG 797  
Qy 276 TrpGlnAlaValProGlyAlaProThrGlyPheIle----- 287  
Db 798 TGGAGCCAGTGGCGGCGCCAGCCGTCCAGCTGGCTCAACAGACACACGGGCGGTTCGCG 857  
Qy 288 -----ProHisLysGlyValPheAspProValAsn 297  
Db 858 GACAAGAACCGCGTGCATCGCGCGCGCCATGAAGTCGTCTCACCCCC----- 911  
Qy 298 HisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAsp 317  
Db 912 AACTTCTCTACGTGACTTACGCGGACTTACCTGCTCCATGGTCCATGGGCGTCACTGGCGAA 971  
Qy 318 ValTrpLysPheSerValThrSerGlyThrTrpThrArgIle----- 331  
Db 972 GTCTGGCGCAGAACCGCAGCTCGGCGCCCTGGGAAGCATTATCTCCCGCGTGGCGAAC 1031  
Qy 332 ---SerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThr 350  
Db 1032 TCGTGGCTGCGCGTACAAACACGAGCGTTCCTCGCGGGCGGATTTGCGGTCTCAGC 1091  
Qy 351 IleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAsp 370  
Db 1092 GTCAGCGGCAACCCCAACCGTCTCGTGTCTCATCAC---CTCGACCGCGACCCCGGA 1148  
Qy 371 ThrIle-----IlePheArgSerThrAspGlyAlaThrTrpThrArgIleTrp 387  
Db 1149 CCCGCCCTCGACAGCATCTACCTCTCAACCGATGCGCGCGGACCTGGAAGACGTCACC 1208  
Qy 388 AspTrpThrSer-----TyrProAsnArgSerLeuArg 398  
Db 1209 CAGTCTGCTCCCGTCCACCTCGAAGGTAACTGGGGCCACCGACTAAACGCGCGCGG 1268  
Qy 399 TyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnPro----- 415  
Db 1269 TAC---AAGAGCGGACGCGCTTCCGTGGCTCGACTTCAACACCGTCCCGAGTGGGG 1325  
Qy 416 -----AsnProValProSerPro-----LysLeuGlyTrpMetAspGluAla 430  
Db 1326 GGATACGCTGGCGCGACGCGTACGCGCGCCGCTCAACAAATTTGGCTGGTGTGATGACGCT 1385  
Qy 431 MetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyr 450  
Db 1386 GTGCTTATCGATCCGTTTCAACCCCGACCATGTATGGCAGCGGCGGACCATCTGG 1445

Qy 451 AlaThrAsnAspLeuThrLys-----TrpAspSerGlyGlyGlnIleHisIle 466  
Db 1446 GCGACCGACAGCTCTCCCGTGTCCAGAAAGACTGG----- 1481  
Qy 467 AlaPro-----MetValLysGlyLeuGluGluThrAlaValAsnAspLeuIle 482  
Db 1482 GCGCGGAGCTGGTACCTCCAGATCGACGGTATCGAGGAGATCGATTCGTGCTCGCTCCG 1541  
Qy 483 SerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHis 502  
Db 1542 TCGCCCAAGAGCGCGCGCTCTCTCGGCGCATCGGTGACATTAGCGGCATGAAGCAC 1601  
Qy 503 AlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThr 522  
Db 1602 GAGCACTCACCAAG---CCCAGAAAGATTTGGTGGCGCCCGTCTTCCACCTCGAC 1658  
Qy 523 SerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspPro 542  
Db 1659 AGCATCGACGCTCGCGGCAACTTCCCAACAGTTGTCTCGCGCGGATCTCTCGGACAC 1718  
Qy 543 SerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrp--- 561  
Db 1719 GAGTACGACAGCGCTGCGCGCGCTGCGTACGCGACTGACGCGGAGACGCGTGGACC 1778  
Qy 562 ---PheGlnGlySerGluProGlyGlyValThrThr-----GlyGlyThrValAla 577  
Db 1779 ATCTTCCCTACTCTGCTGCTGCGCATGAACGCGAGCACCTACCAGGCGCACGATGCA 1838  
Qy 578 AlaSerAlaAspGlySerArgPheValTrpAlaPro-----GlyAspProGlyGlnPro 595  
Db 1839 GTCGACGCGGCGGCGGAGCAGATCTGTGTGTCGCAACCAAGCTTGACGAGGAGGCTCGGA 1898  
Qy 596 ValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAla--- 614  
Db 1899 CCGTGTACTCGCACGACTATGCAAGAGCTGTCT-----GTTCCGCTGGC 1946  
Qy 615 -----AsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyr 629  
Db 1947 GACTTGAAGCCGACACTGCCAATGTCTCTCGACAGGTCAGATGCGACGTTCTAC 2006  
Qy 630 AlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProVal 649  
Db 2007 GCTACCGATGGCGCAAGTTCTTCTGTCGACGCGGCGGGAAGTCGTATGCGCCCAAG 2066  
Qy 650 AlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLys 669  
Db 2067 GCGCGCGGACTTGTCACT-----GSCACATCGCTCATGCTCGCGTGAACCCCTGGGTG 2120  
Qy 670 GluGlyAspLeuTrpLeu---AlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGly 688  
Db 2121 GCGCGGACGCTGCGGTGCTGTCGAGGCGGTCTCTTCCACTCGACCGACTTTGGC 2180  
Qy 689 SerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsnValGlyPheGlyLys 707  
Db 2181 GCCTGTTACAGGAGTAGTACCCCAACGCGACCTCTGTGAGCGTCCGGGCCCCCAAG 2240  
Qy 708 Ser-----AlaProGlySerSerTyrProAlaValPheValValGlyThr--- 722  
Db 2241 TCCAAGTCGACGCGCAAGAGCTAGCGCGCTCCGCGCTCTTCATCTGGGCGACCGAC 2300  
Qy 723 ---IleGlyGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrTrpValLeu 741  
Db 2301 AAGCCTGGAAGCGACATCGGCTGTACCGCTCGACGCAACGCGACACTGAGACCGCG 2360  
Qy 742 IleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAla 761  
Db 2361 GTCAATGACGAGGACCACTACTCTCGGC---CCACCATGATCGAGGCGCGCCCAAG 2417  
Qy 762 AsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle 780  
Db 2418 GTCTACGCGCGCTGTATCTAGGCACGCAACGCGCGGTATCTGTGTACGCGGACCTT 2474



RESULT 11  
US-10-395-241-17  
; Sequence 17, Application US/10395241  
; Publication No. US20040038367A1  
; GENERAL INFORMATION:  
; APPLICANT: YAOI, Katsuro  
; APPLICANT: MITSUISHI, Yaoshi  
; TITLE OF INVENTION: NOVEL XYLOGLUCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE  
; FILE REFERENCE: Q73756  
; CURRENT APPLICATION NUMBER: US/10/395,241  
; CURRENT FILING DATE: 2003-03-25  
; PRIOR APPLICATION NUMBER: JP 2002-83433  
; PRIOR FILING DATE: 2002-03-25  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 17  
; LENGTH: 2481  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Xyloglucan Oligosaccharide-Degrading Enzyme with Histidine Tag  
; NAME/KEY: CDS  
; LOCATION: (1)..(2481)  
US-10-395-241-17

Alignment Scores:  
Pred. No.: 2,38e-69 Length: 2481  
Score: 1124.00 Matches: 298  
Percent Similarity: 48.91% Conservative: 127  
Best Local Similarity: 34.29% Mismatches: 306  
Query Match: 21.89% Indels: 138  
DB: 18 Gaps: 33

US-09-917-376-1 (1-957) x US-10-395-241-17 (1-2481)

QY	52	TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn	71
DB	13	TACGAGTTCAAGAAATGTCGATCGCGCGCGCGGTACATTAACCGGGATTGTCGCGCAC	72
QY	72	GlulGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArgTrpAsp	91
DB	73	CCAAAGACCAAGGACCTGCTGTACCGCGCACGGACATTGGCGCGCGTACCGCTGGGAC	132
QY	92	AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr	111
DB	133	GCAGGCAGTCCAGTGGATCCGCTCAACGACTTATCGAGCGCGCAGGACATGAACATT	192
QY	112	AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal	131
DB	193	ATGGGCACCGAGTCGATCGCGTGGACCCCAACACCCCGCAGCGGTGTACCTCGCGCAG	252
QY	132	GlyMetTyrThr---AsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAsp	150
DB	253	GGCGCGTATGTCGCGCACGAGTGG-----GCGCGGTCTTATGTGTCGGAAGAC	300
QY	151	GlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGly	170
DB	301	CGCGCGCAGTGGTTTCAATCTACGAGTCCGCGTCCCGATCGGCGCCCAACGACATGGGA	360
QY	171	ArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGly	190
DB	361	CGCAACAATGGCGAGCGCTCGCTGTCAACCCGTTCAACTCGAAGCGAGTCTGGATGGGT	420
QY	191	AlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMet	210
DB	421	ACGCGGTACA---GAGGGTATCTGGAGAGTTCGGACCGCGCCCAAGACCTGGCAACAGTC	477
QY	211	ThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGln	230
DB	478	ACGTCCATCCCGACCGCTTC-----ACCAACGGTATCGGATACAG	519
QY	231	SerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGln	250
DB	520	TGC-----GTCAATTTTCGACCCC-----GAA	540
QY	251	AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArg	270
DB	541	CCTAATGGCACCATTCTACGCGAGCGACTGCCCCCGCAGGCG---ATGTACGTACACGAC	597
QY	271	AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIle-----	287
DB	598	GACGCGGTGTCTCGTGGAGCAGTGGCGGCCAGCCGTCACAGTGGCTCAACAGGACC	657
QY	288	-----ProHisLysGlyVal	292
DB	658	ACGGCGCGGTTCCTCCGACAAAGCCCGCTGATCGCGCGCAGCCCATGAAGTCGCT	717
QY	293	PheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAsp	312
DB	718	CTCACCCCC-----AACTTCTCTACGTACTTACGCGGACTACCTGGTGTCCATGGGCG	771
QY	313	GlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArgIle---	331
DB	772	GTACGTTTCGCGAAAGTCTGGCGCCAGAACCGCACCTCGGCGCGCTCGGAGCATTA	831
QY	332	-----SerProValProSerThrAspThrAlaAsnAspTyrPheGly	345
DB	832	CCCCCGGTCCGCAACTCTGCTGCTGCCCGGTACAAACACAGACGTTCCCTCGCGGCGGA	891
QY	346	TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle	365
DB	892	TTTTTGGGTCTCAGCTCGACGCGACCAACCCCAACCGTCTCGTCTCATCACC---CTC	948
QY	366	SerTrpTrpProAspThrIle-----IlePheArgSerThrAspGlyGlyAlaThr	382
DB	949	GACCGCGACCCCGGACCCGCTCGACAGCATCTACTCTCAACCGGATCGCGCGGAC	1008
QY	383	TrpThrArgIleTrpAspTrpThrSer-----TyrPro	393
DB	1009	TGGAAGGAGCTACCCAGCTCTCGTCCCGTCCAACTCAAGTAATACTGGGCGCACCG	1068
QY	394	AsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyVal	413
DB	1069	ACTAACCGCGCGGTAC---AAGGACGCGACGCTCTTCCGTGGCTCGACTTCAACAC	1125
QY	414	GlnPro-----AsnProValProSerPro-----LysLeuGly	425
DB	1126	GGTCCCGACGTTGGGGGATACGCGTGGCGCGCAGTACGCGCGGCTCAACCAAGTTGGC	1185
QY	426	TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr	445
DB	1186	TGGTGGATGAGCGCTGTGCTTATCGATCCGTTCAACCCCGCAGCCTGTGTAGTACG	1245
QY	446	GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLys-----TrpAspSerGly	461
DB	1246	GGGCGCACCATCTGGGCGACCGACACGCTCTCCGCTCGAGAGGACTGG-----	1296
QY	462	GlyGlnIleHisIleAlaPro-----MetValLysGlyLeuGluGluThrAla	477
DB	1297	-----GCGCGGAGTGGTGTCTCCAGATCGACGCTATCGAGGAGATGCG	1341
QY	478	ValAsnAspLeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeu	497
DB	1342	ATCTCTGCTCGCTCGCGTCCGCAAGAGCGCGCGCTCTCTGTCGGGCATCGGTACATT	1401
QY	498	GlyGlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProVal	517
DB	1402	AGCGGATGAGCAGCAGCAGCTCACCAAG---CCCCAGAGATGTTGGTGGCGCCCGAC	1458
QY	518	PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla	537
DB	1459	TTCTCCAACTCGACAGCATCGACGCTCGCGCAACTTCCCGCAACGTTGTCGCGCGCC	1518
QY	538	GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly	557











## Alignment Scores:

Pred. No.: 4,68e-56 Length: 5698  
 Score: 940.00 Matches: 374  
 Percent Similarity: 30.11% Conservative: 136  
 Best Local Similarity: 22.08% Mismatches: 315  
 Query Match: 18.31% Indels: 871  
 DB: 18 Gaps: 59

US-09-917-376-1 (1-957) x US-10-420-191-1 (1-5698)

QY 25 SerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAla 44  
 DB 128 TCTCGAGTCCTTCTTCCAGAGAGCTCAGGAACGGGAACAGGACCCCGCAGTAGGACGG 169  
 QY 45 HisAlaIleThrGln----- 50  
 DB 170 CATGCTGCTTCTTCCAGAGAGCTCAGGAACGGGAACAGGACCCCGCAGTAGGACGG 229  
 QY 51 -----ProTyrThrTrpSerAsnValAlaIle---GlyGlyGlyGlyPheValAspGly 67  
 DB 230 GTACGACGGAATTCATGGAAGACGTCAGCTCGGCGCGCGGCTTCGTCCCGCGC 289  
 QY 68 IleValPheAsn-----GluGlyAlaPro----- 75  
 DB 290 ATCATCTTCCAAAGTACCTCTTCAGTTCGAGCGCGCGCGCGCGCGAGCGGCGC 349  
 QY 76 -----GlyIleLeuTyrValArgThrAspIleGlyGlyMet 87  
 DB 350 TAGTAGAAGGTTCCCAAGACAAAGCGCTAGCATATGCACGAACAGATATTGGCGGCTG 409  
 QY 88 TyrArgTrp----- 90  
 DB 410 TACCGCTCAAAGGTTCTGTTTCCGATCGTATACGTGTTGCTATACCCCGCGAC 469  
 QY 91 -----AspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpVal----- 104  
 DB 470 ATGGCGGAGTTCGCGAGCATCTAGTACCGCGCGCGCGCGCGCGTACCGATGGATGCTGATAATGCC 529  
 QY 105 GlyTrp-----AsnAsnTrpGly----- 110  
 DB 530 GGCTGGCAACAGCGGCTGCTGAGTACCTGGCGGCGAGTGCCTACCGTACGACTATTACGG 589  
 QY 111 -----TyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysVal 127  
 DB 590 CCGACCGGTGTTCTGGGGCATCGACGCTGTGGCTTGATCCGCGAGGACGACAAAGGTG 649  
 QY 128 TrpAlaAlaVal----- 131  
 DB 650 TATGCCGAGTGATACCCGCTAGTGGCAACACGCGNACTAGGGGCTCTGCTAGTTTCCAC 709  
 QY 132 -----GlyMetTyrThrAsnSerTrpAsp----- 139  
 DB 710 ATACGGCTCACGGCATGTATACGAACAGCTGGTC-TGTGATGTCTCTCAGATCTAGACC 768  
 QY 139 ----- 139  
 DB 769 TATGATTGGACGGCGGTATATGCTTGTTCGACCAACAGACACTACAGAGAGTCTAGATCTGG 828  
 QY 140 -----ProAsnAspGlyValAlaIleLeuArg 147  
 DB 829 ATACTAAGCTGCTGACATTTGGCCATATATAGGGATCCGAGTATAGGAGCCATCATTCGC 888  
 QY 148 SerSerAspGln----- 151  
 DB 889 TCGTCAGACCGGACTGTAAACGGGTATATATCCCTAGGCTCATTTACCTCGGTAGTAAGCG 948  
 QY 152 -----GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsn 167  
 DB 949 AGCAGCTGGCGGGCGCAACGTGGTCTTCCACCACTTGCCCTTCAAAGTCGGGGGTAAAC 1008  
 QY 168 MetProGlyArg----- 171

DB 1009 ATGCCAGGACGCCCGCGTTCACACCGAAGTGTTGAAACGGGAAGTTCAGCCCCCATTTG 1068  
 QY 172 -----GlyMetGlyGluAlaValAspProAsnAsnAspAsnIleLeu 187  
 DB 1069 TACGGTCTCGGGAGCCGGAGAGCGTCTGGCTGTGATCCGGCCAACTCCAAACATCATC 1128  
 QY 188 TyrPheGlyAlaPro----- 192  
 DB 1129 TACTTTGGTCTCTCTCGGCCTCTCGCAGACCGACAGCTAGGCGGTTGAGGTGTAGTAG 1188  
 QY 193 -----SerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThrTrp 207  
 DB 1189 ATGAAACACGACGCTCAGAAACCGCCTCTGAAAGTCTACGACGGCGCGTACCTTT 1248  
 QY 208 SerGlnMet----- 210  
 DB 1249 TCCAAAGTCTCGCGAGTCTTTTGGCGGAGACCTTCAGATGCTGCCCGCGCACTCGAAA 1308  
 QY 211 -----ThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrTrp 227  
 DB 1309 AGGTTCCAGAGCTCGTTTCCAGGCAACTGGGACGTATCCAGACCCGAGTGATTCCAAC 1368  
 QY 228 GlyTyrGln----- 230  
 DB 1369 GGCTACAAACAGCAGCAGCAAGTGCCTGACCTGCATGTAGGGTCTGGGCTCACTAAGGTTG 1428  
 QY 231 -----SerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSer 247  
 DB 1429 CCGATGTTGTTCGACAAAGGACTCATGTGGGTACGTTTCGACTCAACAGCAGCAGC 1488  
 QY 248 LeuGlyGlnAla----- 251  
 DB 1489 ACCGGGGAGCCCTGTTCTGCTGAGTACACCAATGCAAGCTGAGTTGGTCTGTCGTGC 1548  
 QY 252 -----SerLysThrIlePheValGlyValAlaAspProAsnAsnProValPhe 267  
 DB 1549 TGGCCCCCTCGGACGCTCTCGTATCTTTGTTGGCAGCGCTGATAACATCATCTGCTCAGTC 1608  
 QY 268 Trp----- 268  
 DB 1609 TATGTGACACGTGCAGAGCATAGAAAACACCGTCCGACTATTGTAGTGACGAAGTCAG 1668  
 QY 269 -----SerArgAspGlyAlaValThrTrpGlnAlaValProGlyAlaProThrGlyPhe 286  
 DB 1669 ATACACTCGTCAATGCCGCTCCAGCTGAGTGTGTACCGGGGCGCAGCCAGGGAATAC 1728  
 QY 287 IleProHis----- 289  
 DB 1729 TTTCTCTCAAGTTACGGCGGAGGTGCACCTCAGCAGATGGCCCGCTCGCTCCCTTTATG 1788  
 QY 290 LysGlyValPhe-----AspProValAsnHisValLeuTyrIleAlaThrSerAsn 306  
 DB 1789 AAAGAGGTGTTTCGCGAAACTGCAGCCAGCAGAGAGGCGCTTGTATCTGACCTATTCCGAT 1848  
 QY 307 ThrGlyGlyPro----- 310  
 DB 1849 GGCAAGGCGCGCGCTTTGACGTCGTCTCTTCGGAACATAGACTGGATAAGGCTA 1908  
 QY 311 -----TyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGly 326  
 DB 1909 CGGTGTCGGGTATGATGGCACACTTGGCTCAGTGTGAGGTACGACATTCGAGGGGA 1968  
 QY 327 ThrTrpThrArgIleSerProValProSerThrAspThr----- 339  
 DB 1969 ACTTGGAAAGACATACACTACCGTGTGAACCGAGTCAACCTCCATGCTGTAAACGTCGCCCT 2028  
 QY 340 -----AlaAsnAspTyrPheGlyTyrSerGlyLeu 349  
 DB 2029 TGAACCTTTCTGATCACCCCTGTCTCTGGATCAGATCTATCTTTTGGCGCCTT 2088  
 QY 350 ThrIleAsp----- 352  
 DB 2089 GGCCTCGATTGTAGTGGGACAGAGACCTAGTCTAGATATGAAACCGAAACCGCCGGA 2148











```
QY 649 -----ValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaVal 666
Db 979 CGGTCGTGCGCGCTGGCTTCCAGCAGC-----ATCGGCATGCTCATCGACACCTTA 1032
QY 667 ProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsn 686
Db 1033 -----CGCAAC 1038
QY 687 GlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnVal----- 703
Db 1039 GGTGGGTGTCGCAACGCAACCAAGCCGAGCAGCCGCGGACCGCATGTCAACACCTTC 1098
QY 704 -----Gly 704
Db 1099 GTCACACGATCGAAGATTGACCTTCGGCGAGCAGCCGCGCCTGTGGTGAACAGCAACGCT 1158
QY 705 PheGlyLysSerAlaProGlySerSerTyrProAlaValPhe----- 718
Db 1159 GCGGCTCGCCAGCCGCGCAGCAGCCGAGCGACTTCCGCAACGCGCACCTCGAC 1218
QY 719 -----ValValGlyThrIleGlyValThrGlyAla----- 729
Db 1219 GCGTATGCTCGATCAAGCCGCGGTGAGTCGACGCGCACCAAGCGCTGCGAGCGATCCG 1278
QY 730 -----TyrArgSerAspAsp---CysGlyThrThrTrpValLeuIleAsnAspAsp 745
Db 1279 ACAACTGCAAGAAGTGGACCCCATGTGCGACCGCGATACAGCAGC----- 1326
QY 746 GlnHisGlnTyrGlyAsnTrpGlyGlnAla-----IleThrGlyAspHis 760
Db 1327 -----TCGTACGGGGTACTGACCAACCGCTTACCGAACTCCCGCATCGCGCCAGTGG 1380
QY 761 AlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle 780
Db 1381 TTCCCGGCGAGTTTGACCACTGTGTCGCAAGCAGCGCCAGCGGTGCCGACGTCGACC 1440
QY 781 GlyGly-----AlaProSerGlySerProSerProSerValSer 793
Db 1441 AGCTCGAGCCGCGCCTCCGCGCGAGTCCGTGCGGTTCGCGGAGTCCGAGCCCGAGT 1500
QY 794 ProSerAlaSerProSerProSerProSerProSerProSerProSerProSer 813
Db 1501 CCGAGCCGAGAGTCGCGCTGCGCGCTCTCCGAGCTCGAGCCCGTCTCCGTCG 1560
QY 814 Pro-----SerProSerSerProSerProSerProSerProSer 825
Db 1561 CCGAGCCGAGTCCGAGCCGAGTAGCTCGCGTCTCGCGTCTCCGAGCTCGAGCCCGTCT 1620
QY 826 ProSerProSerProSerProSerProSerProSerProSerProSerProSer 845
Db 1621 CCGTCCGCGAGCCCGAGTCCGAGCCGAGTAGCTCGCGTCTCGCGTCTCCGAGCTCGAGC 1680
QY 846 ProSerSerProSerProSerProSerProSerProSerProSerProSerProSer 865
Db 1681 CCGTCTCGCTCGCGAGCCGAGTCCGAGCCGAGTAGCTCGCGTCTCGCGTCTCCGAGCTCGAGC 1740
QY 866 SerSerProValSerGlyGlyValGlnTyrLysAsnAsnAspSerAlaProGly 885
Db 1741 TCGTCGCGGTGTGCGGTGGGCTGAAGGTGAGTACAGCAACATGATTCGCGCGCGGT 1800
QY 886 AspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySerSerSerValAspLeu 905
Db 1801 GATAACCAAGATCAACACCGGCTCTCCAGTTGGTGAATACCGGCTCGCTCGGTGGATTG 1860
QY 906 SerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrIleValTyr 925
Db 1861 TCGACGGTGACGTGCGGTACTGGTTTACCCCGGGATGGTGGGTCTCGACACTGGGTGATC 1920
QY 926 AsnCysAspTrpAlaAlaIleGlyCysGlyAsnIleArgAlaSerPheGlySerValAsn 945
Db 1921 AACTGTGACTGGGCGCGATGGGGTGTGGGAATATCCGCGCCTCTGTTGGCTCGGTGAAC 1980
```

```
QY 946 ProAlaThrProThrAlaAspThrTyrLeuGln 956
Db 1981 CGGCGACGCGCGACGCGGACACCTACTCTGCAG 2013
RESULT 15
US-09-917-383-2
; Sequence 2, Application US/09917383
; Publication No. US20030104522A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,383
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 3687
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-383-2
Alignment Scores:
Pred. No.: 4,86e-48 Length: 3687
Score: 823.00 Matches: 251
Percent Similarity: 47.07% Conservative: 46
Best Local Similarity: 39.78% Mismatches: 173
Query Match: 16.03% Indels: 164
DB: 10 Gaps: 24
US-09-917-376-1 (1-957) x US-09-917-383-2 (1-3687)
```

```
QY 426 TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr 445
Db 307 TGGATGGACCGCATCGCTCGCATCAACGGCGTCAACGGC-----GGACCC 351
QY 446 GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHis 465
Db 352 GCGTTGACGACATATCTGGACGCGCGCTCTCCAG-----CAGCAGGA 396
QY 466 IleAlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProPro 485
Db 397 ACCACCCCTGAAGTC-----ATTGATTTGTCACTACGATCTG-----CCGGGA 441
QY 486 SerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyPheThrHisAlaAspVal 505
Db 442 CCGGACTGCGCGCGCTCGCTCCCAACGCGGAACTGCCGCT---ACGCGACGAGTTTG 498
QY 506 ThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAsp 525
Db 499 CAGACCTAT---GAAACGCGAGTACATCGATTCGAGTCTCCGAGTATCTCCGAGC 555
QY 526 TyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerGln 545
Db 556 TACTCCAGCTCGGATCGGTGACGATCAT---GAGCCGAGCTCGCTG 600
QY 546 ProAsnAspArgHisValAlaPheSerThrAspGlyLysAsnTrpPheGlnGlySer 565
Db 601 CCNAACGC---GGTCAC-----CAATAT 620
QY 566 GluProGlyGlyValThrThrGlyThrValAla-----AlaSerAlaAspGly 582
Db 621 GAGCATTCAGCGTGTGCAACGCGCGTTCGCTATT-ACGAGCAAGGATCGAGTACGCGC 679
QY 583 SerArg-----PheValTrpAlaProGlyAspPro--- 592
```







GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 07:48:48 ; Search time 141.761 Seconds  
(without alignments)  
2018.911 Million cell updates/sec

Title: US-09-917-376-3  
Perfect score: 4036  
Sequence: 1 ATQPTWNSVAIGGGFVD.....YIGTNGRIVYGDIGGAPSG 740

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_16Dec04.\*  
1: Geneseqp1980s.\*  
2: Geneseqp1990s.\*  
3: Geneseqp2000s.\*  
4: Geneseqp2001s.\*  
5: Geneseqp2002s.\*  
6: Geneseqp2003as.\*  
7: Geneseqp2003bs.\*  
8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4036	100.0	740	6	Abp73016 Amino aci
2	4036	100.0	740	7	Add22926 Acidother
3	4036	100.0	740	7	Add22923 Acidother
4	4036	100.0	957	6	Abp73015 Amino aci
5	4036	100.0	957	7	Add22921 Acidother
6	4036	100.0	957	8	Ado52313 Acidother
7	3920.5	97.1	733	8	Ado52315 A. cellu
8	2063	51.1	940	6	Abp99489 Amino aci
9	1680	41.6	726	7	Add22927 Aspergill
10	1579	39.1	838	7	Add42055 Trichoder
11	1557.5	38.6	818	8	Adh51581 Trichoder
12	1117	27.7	789	8	Adh19120 Geotrichu
13	1117	27.7	812	8	Adh19118 Geotrichu
14	1114	27.6	826	8	Adh19124 Geotrichu
15	979	24.3	756	8	Adr90298 Geotrichu
16	979	24.3	757	8	Adr90302 Geotrichu
17	979	24.3	776	8	Adr90296 Geotrichu
18	841.5	20.8	555	7	Add424919 Xanthomon
19	238	5.9	2468	6	Abu38411 Protein e
20	238	5.9	2468	6	Abp59933 Microbia
21	219.5	5.4	2736	7	Abp81481 Pseudomon
22	199.5	4.9	2435	6	Abu19529 Protein e
23	193	4.8	1465	8	Adr24922 Bacterial
24	181.5	4.5	3892	6	Ada34216 Acinetoba
25	181	4.5	1308	6	Abu16642 Protein e

ALIGNMENTS

RESULT 1

ABP73016

ID ABP73016 standard; peptide; 740 AA.

XX

AC ABP73016;

XX

DT 03-JUN-2003 (first entry)

XX

DE Amino acid sequence of the avicelase Aviii catalytic domain.

XX

KW Avicelase; Aviii; glycoside hydrolase; enzyme; cellulase; biofuel;  
KW detergent; pulp processing; paper processing; feed processing; textile;  
KW cellulose.

XX

OS Acidothermus cellulolyticus.

XX

PN WO2003012090-A2.

XX

PD 13-FEB-2003.

XX

PF 28-JUL-2001; 2001WO-US023818.

XX

PR 28-JUL-2001; 2001WO-US023818.

XX

PA (WIDE ) MIDWEST RES INST.

XX

PI Ding S, Adney WS, Vinzant TB, Himmel MB;

XX

DR WPI; 2003-248177/24.

XX

PT New thermostable Aviii peptide from Acidothermus cellulolyticus, useful  
PT for degradation of cellulose or in generating anti-Aviii antibodies for  
PT purifying recombinant Aviii polypeptides from genetically engineered  
PT host cells.

XX

PS Claim 6; Page 8; 44pp; English.

XX

CC The present sequence is derived from a thermostable avicelase, designated  
CC Aviii. Aviii is a member of the glycoside hydrolase family of enzymes,  
CC and is a cellulase. Aviii is useful in the conversion of biomass to  
CC biofuels and biofuel additives. It may be useful in the production of  
CC detergents, pulp and paper processing, food and feed processing and in  
CC textile processes. The thermostable Aviii peptide is useful in the  
CC degradation of cellulose, and in generating recombinant Aviii polypeptides  
CC antibodies that are useful in purifying recombinant Aviii polypeptides  
CC from genetically engineered host cells, in detecting Aviii polypeptide  
CC expression as well as a reagent tool for characterizing the molecular  
CC actions of the polypeptide. The Aviii polynucleotide is useful as a







421 APWKLEETAVNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIETSPVFTTGTSDY 480  
 481 AELNPSIIIVRAGSFDPSQPNDRHVAFTDGGKWFQSGEPGGVTTGGTVAASADGSRFV 540  
 481 AELNPSIIIVRAGSFDPSQPNDRHVAFTDGGKWFQSGEPGGVTTGGTVAASADGSRFV 540  
 541 WAPGDPQPVVYAVFGNSWAASQGVPAQAQIRSDRVNPKTFYALSNGTFFYRSTDDGGVTF 600  
 541 WAPGDPQPVVYAVFGNSWAASQGVPAQAQIRSDRVNPKTFYALSNGTFFYRSTDDGGVTF 600  
 601 QPVAAGLPSSGAVGMFHAHPGKEGDLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 660  
 601 QPVAAGLPSSGAVGMFHAHPGKEGDLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 660  
 661 KSAFGSSYPVAVVGTGGVTVGAYRSDDCGTTWVLINDDQHQYGNWQAITGDHANLRRV 720  
 661 KSAFGSSYPVAVVGTGGVTVGAYRSDDCGTTWVLINDDQHQYGNWQAITGDHANLRRV 720  
 721 YIGTNGRGIVYDGGAPSG 740  
 721 YIGTNGRGIVYDGGAPSG 740

RESULT 3

ADD22923 ID ADD22923 standard; protein; 740 AA.

AC ADD22923;

XX 15-JAN-2004 (first entry)

DT Acidothermus cellulolyticus avicelase AvIII catalytic domain.

DE enzyme; AvIII; cellulose reduction; agricultural biomass;  
 KW municipal solid waste; glycoside hydrolase; avicelase.

XX Acidothermus cellulolyticus.

OS US2003108988-A1.

PN 12-JUN-2003.

XX 18-OCT-2002; 2002US-00155400.

XX 28-JUL-2001; 2001US-00917376.

PR (DING/) DING S.

PA (ADNEY/) ADNEY W S.

PA (VINZ/) VINZANT T B.

XX (HIMM/) HIMMEL M E.

PI Ding S, Adney WS, Vinzant TB, Himmel ME;

XX WPI; 2003-810853/76.

XX New isolated thermal tolerant avicelase polynucleotide useful for  
 PT detection of a polynucleotide encoding AvIII and for reducing cellulose  
 PT in a starting material, e.g. municipal solid waste.

PS Claim 16; SEQ ID NO 3; 29pp; English.

XX The invention relates to an isolated polynucleotide molecule encoding a  
 CC thermostable AvIII polypeptide. The polynucleotide is useful for  
 CC detection of a polynucleotide encoding AvIII. The polynucleotide is  
 CC useful for reducing cellulose in a starting material which involves  
 CC administering to the starting material, e.g. agricultural biomass or  
 CC municipal solid waste, a polypeptide molecule of the polynucleotide. The  
 CC method further comprises administering a second polypeptide molecule  
 CC chosen from the glycoside hydrolase family of proteins. The present  
 CC sequence represents the amino acid sequence of Acidothermus  
 CC cellulolyticus avicelase AvIII catalytic domain.

XX Sequence 740 AA;

OS Acidothermus cellulolyticus.

Query Match 100.0%; Score 4036; DB 7; Length 740;  
 Best Local Similarity 100.0%; Pred. No. 8.4e-271;  
 Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATTQPTWNSNVAIGGGGFDGIVFNEGAPGILYVTRDIGGMYRDAANGRWIPLLDWVGW 60  
 DB 1 ATTQPTWNSNVAIGGGGFDGIVFNEGAPGILYVTRDIGGMYRDAANGRWIPLLDWVGW 60  
 QY 61 NNWGYGVVSIADPINTNKWAAVGMVYNSNDPNDGAILRSSDQAGATWITPLPFLKGG 120  
 DB 61 NNWGYGVVSIADPINTNKWAAVGMVYNSNDPNDGAILRSSDQAGATWITPLPFLKGG 120  
 QY 121 NNWPGMGRLAVDPNNDNILYFGAPSGKGLWRSTDGATWSQMTNPPDVTYIANPTDT 180  
 DB 121 NNWPGMGRLAVDPNNDNILYFGAPSGKGLWRSTDGATWSQMTNPPDVTYIANPTDT 180  
 QY 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIFGVADPNNPVFWRDGGATWQAVGAPTGF 240  
 DB 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIFGVADPNNPVFWRDGGATWQAVGAPTGF 240  
 QY 241 IPHKGVPDPVNHVLYIATSNCTGGPYDSSGDKWKFVSTGVTWTRISVPSTDTANDYFGY 300  
 DB 241 IPHKGVPDPVNHVLYIATSNCTGGPYDSSGDKWKFVSTGVTWTRISVPSTDTANDYFGY 300  
 QY 301 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDCGATWTRIDWTSYPNRSRLYVLDISAE 360  
 DB 301 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDCGATWTRIDWTSYPNRSRLYVLDISAE 360  
 QY 361 PMLTFGVQPNPPVPSPKLGWMDMAIDPFNSDRMLYGTGATLYATNDLTWDSGGQIHI 420  
 DB 361 PMLTFGVQPNPPVPSPKLGWMDMAIDPFNSDRMLYGTGATLYATNDLTWDSGGQIHI 420  
 QY 421 APWKLEETAVNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIETSPVFTTGTSDY 480  
 DB 421 APWKLEETAVNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIETSPVFTTGTSDY 480  
 QY 481 AELNPSIIIVRAGSFDPSQPNDRHVAFTDGGKWFQSGEPGGVTTGGTVAASADGSRFV 540  
 DB 481 AELNPSIIIVRAGSFDPSQPNDRHVAFTDGGKWFQSGEPGGVTTGGTVAASADGSRFV 540  
 QY 541 WAPGDPQPVVYAVFGNSWAASQGVPAQAQIRSDRVNPKTFYALSNGTFFYRSTDDGGVTF 600  
 DB 541 WAPGDPQPVVYAVFGNSWAASQGVPAQAQIRSDRVNPKTFYALSNGTFFYRSTDDGGVTF 600  
 QY 601 QPVAAGLPSSGAVGMFHAHPGKEGDLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 660  
 DB 601 QPVAAGLPSSGAVGMFHAHPGKEGDLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 660  
 QY 661 KSAFGSSYPVAVVGTGGVTVGAYRSDDCGTTWVLINDDQHQYGNWQAITGDHANLRRV 720  
 DB 661 KSAFGSSYPVAVVGTGGVTVGAYRSDDCGTTWVLINDDQHQYGNWQAITGDHANLRRV 720  
 QY 721 YIGTNGRGIVYDGGAPSG 740  
 DB 721 YIGTNGRGIVYDGGAPSG 740

RESULT 4

ABP73015

ID ABP73015 standard; protein; 957 AA.

XX AC ABP73015;

XX DT 03-JUN-2003 (first entry)

XX DE Amino acid sequence of the avicelase AvIII.

XX KW Avicelase; AvIII; glycoside hydrolase; enzyme; cellulase; biofuel;  
 KW detergent; pulp processing; paper processing; feed processing; textile;  
 KW cellulose.  
 XX OS Acidothermus cellulolyticus.



XX	Key	Location/Qualifiers	
FH	Misc-difference 957	/note= "unspecified residue encoded by N"	
FT			
XX			
XX	W02003012090-A2.		
XX	13-FEB-2003.		
XX			
XX	28-JUL-2001; 2001WO-US023818.		
XX	28-JUL-2001; 2001WO-US023818.		
XX	(MIDE ) MIDWEST RES INST.		
XX			
XX	Ding S, Adney WS, Vinzant TB, Himmel ME;		
XX			
XX	WPI; 2003-248177/24.		
DR	N-PSDB; ABZ77632.		
XX			
XX	New thermostable AviiiI peptide from Acidothermus cellulolyticus, useful for degradation of cellulose or in generating anti-AviiiI antibodies for purifying recombinant AviiiI polypeptides from genetically engineered host Cells.		
XX			
PS	Claim 2; Page 20; 44pp; English.		
XX			
CC	The present sequence represents a thermostable avicelase polypeptide, designated AviiiI. AviiiI is a member of the glycoside hydrolase family of enzymes, and is a cellulase. AviiiI is useful in the conversion of biomass to biofuels and biofuel additives. It may be useful in the production of detergents, pulp and paper processing, food and feed processing and in textile processes. The thermostable AviiiI peptide is useful in the degradation of cellulose, and in generating specific anti-AviiiI antibodies that are useful in purifying recombinant AviiiI polypeptides from genetically engineered host cells, in detecting AviiiI polypeptide expression, as well as a reagent tool for characterizing the molecular actions of the polypeptide. The AviiiI polynucleotide is useful as a source of probes or primers in various diagnostic assays		
XX			
SQ	Sequence 957 AA;		
	Query Match	100.0%; Score 4036; DB 6; Length 957;	
	Best Local Similarity	100.0%; Pred. No. 1.2e-270;	
	Matches 740; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60	
Db	47	ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 106	
QY	61	NNWGYGVVSIADPINTNKVAAVGMVYNSWDNDGAILRSSDQGTWQITPLPFKLG 120	
Db	107	NNWGYGVVSIADPINTNKVAAVGMVYNSWDNDGAILRSSDQGTWQITPLPFKLG 166	
QY	121	NMPCRGGERLAVDPNNDNIIYFCAPSGKGLWRSTDGATWSQMTNPDVGTIYANPTDT 180	
Db	167	NMPCRGGERLAVDPNNDNIIYFCAPSGKGLWRSTDGATWSQMTNPDVGTIYANPTDT 226	
QY	181	TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVDAPNNPVFWSRDGGATWQAVFGAPTGF 240	
Db	227	TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVDAPNNPVFWSRDGGATWQAVFGAPTGF 286	
QY	241	IPHKGVDPVNHVLIYATSNTPGPGYDSSGVWKFVSVTSGTWTTRISVPSTDTTANDYFGY 300	
Db	287	IPHKGVDPVNHVLIYATSNTPGPGYDSSGVWKFVSVTSGTWTTRISVPSTDTTANDYFGY 346	
QY	301	SGLTIDRQHNTIMVATQISWMPDTIIFRSDGATWTRIDWTSYNRSLRYVLDISAE 360	
Db	347	SGLTIDRQHNTIMVATQISWMPDTIIFRSDGATWTRIDWTSYNRSLRYVLDISAE 406	
QY	361	PWLTFGVQPNPVPSPKLGWDEAMADFPNSDRMLYGTATLYATNDLTKWDSGGQIHI 420	
Db	407	PWLTFGVQPNPVPSPKLGWDEAMADFPNSDRMLYGTATLYATNDLTKWDSGGQIHI 466	
QY	421	APMVKGLEETA VNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTGSVDY 480	
Db	467	APMVKGLEETA VNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTGSVDY 526	
QY	481	AELNPSIIVRAGSFDPSQPNDRHVAFSTDGKQWFFQSGEPGGVTTGGTVAASADGSRFV 540	
Db	527	AELNPSIIVRAGSFDPSQPNDRHVAFSTDGKQWFFQSGEPGGVTTGGTVAASADGSRFV 586	
QY	541	WAPGDGQPVVYAVGPGNSWAASQGVPANAOIRSDRVNPKTFYALSNGTFRSTDCGVTF 600	
Db	587	WAPGDGQPVVYAVGPGNSWAASQGVPANAOIRSDRVNPKTFYALSNGTFRSTDCGVTF 646	
QY	601	QPVAAGLPSSGAVGMFHAVPCKEGDMLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 660	
Db	647	QPVAAGLPSSGAVGMFHAVPCKEGDMLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 706	
QY	661	KSAPGSSYPVAVVGTIGVGTGAYRSDDCGTTWVLINDDQHOYGNWQAITGDHANLRV 720	
Db	707	KSAPGSSYPVAVVGTIGVGTGAYRSDDCGTTWVLINDDQHOYGNWQAITGDHANLRV 766	
QY	721	YIGTNGRGIVYGDIGGAPSG 740	
Db	767	YIGTNGRGIVYGDIGGAPSG 786	
XX	RESULT 5		
XX	ADD22921		
ID	ADD22921 standard; protein; 957 AA.		
XX	AC	ADD22921;	
XX	DT	15-JAN-2004 (first entry)	
XX	DE	Acidothermus cellulolyticus avicelase AviiiI.	
XX	KW	enzyme; AviiiI; cellulose reduction; agricultural biomass; municipal solid waste; glycoside hydrolase; avicelase.	
XX	OS	Acidothermus cellulolyticus.	
FH	Key	Location/Qualifiers	
FT	Misc-difference 957	/label= Unknown	
FT		/note= "Encoded by N"	
XX			
XX	US2003108988-A1.		
XX	12-JUN-2003.		
XX			
PF	18-OCT-2002; 2002US-00155400.		
XX			
PR	28-JUL-2001; 2001US-00917376.		
XX			
PA	(DING/) DING S.		
PA	(ADNE/) ADNEY W S.		
PA	(VINZ/) VINZANT T B.		
PA	(HIMM/) HIMMEL M E.		
XX			
PI	Ding S, Adney WS, Vinzant TB, Himmel ME;		
XX			
DR	WPI; 2003-810853/76.		
DR	N-PSDB; ADD22922.		
XX			
PT	New isolated thermal tolerant avicelase polynucleotide useful for detection of a polynucleotide encoding AviiiI and for reducing cellulose in a starting material, e.g. municipal solid waste.		
XX			
PS	Claim 16; SEQ ID NO 1; 29pp; English.		
XX			
CC	The invention relates to an isolated polynucleotide molecule encoding a thermostable AviiiI polypeptide. The polynucleotide is useful for detection of a polynucleotide encoding AviiiI. The polynucleotide is		



CC useful for reducing cellulose in a starting material which involves  
CC administering to the starting material, e.g. agricultural biomass or  
CC municipal solid waste, a polypeptide molecule of the polynucleotide. The  
CC method further comprises administering a second polypeptide molecule  
CC chosen from the glycoside hydrolase family of proteins. The present  
CC sequence represents the amino acid sequence of Acidothermus  
CC cellulolyticus avicelase Aviii.

XX SQ Sequence 957 AA;  
Query Match 100.0%; Score 4036; DB 7; Length 957;  
Best Local Similarity 100.0%; Pred. No. 1.2e-270;  
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATTQPTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60  
DB 47 ATTQPTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 106  
QY 61 NNMWYGVVSIADPINTNKVAAVGMVYTNWDNDGAILRSSDQATWQITPLPKLGG 120  
DB 107 NNMWYGVVSIADPINTNKVAAVGMVYTNWDNDGAILRSSDQATWQITPLPKLGG 166  
QY 121 NMPGRGMRGLAVDPNNNDILYFGAPSGKGLWRSSTDGATWQMTNPDVGYIANTPTD 180  
DB 167 NMPGRGMRGLAVDPNNNDILYFGAPSGKGLWRSSTDGATWQMTNPDVGYIANTPTD 226  
QY 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNPNFVWSRDGGATWQVPGAPTGF 240  
DB 227 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNPNFVWSRDGGATWQVPGAPTGF 286  
QY 241 IPHKGVPDNNHVLVIATNTGPGVDSGDVWKFVSVTSGTWTRISPVSTDTANDYFY 300  
DB 287 IPHKGVPDNNHVLVIATNTGPGVDSGDVWKFVSVTSGTWTRISPVSTDTANDYFY 346  
QY 301 SGLTDRQHPNTIMVATQISWPPDTIIFRSTGGATWTRIMDWTSPNRSRLRVLDISAE 360  
DB 347 SGLTDRQHPNTIMVATQISWPPDTIIFRSTGGATWTRIMDWTSPNRSRLRVLDISAE 406  
QY 361 PMLTFGVQNPVPSPKLGWDEMAIDPFNSDRMLYGTGATLYATNDLTKWDGSGQIHI 420  
DB 407 PMLTFGVQNPVPSPKLGWDEMAIDPFNSDRMLYGTGATLYATNDLTKWDGSGQIHI 466  
QY 421 APMVKGLETA VNDLISPPSGAPLISALGDLGFTHADVAVPSTIPTSFPVFTGTSVDY 480  
DB 467 APMVKGLETA VNDLISPPSGAPLISALGDLGFTHADVAVPSTIPTSFPVFTGTSVDY 526  
QY 481 AELNPSIIVRAGSFDPSQPNDRHVAFSTDGKKNWFGQSEPGVTTGGTVAASADGRFV 540  
DB 527 AELNPSIIVRAGSFDPSQPNDRHVAFSTDGKKNWFGQSEPGVTTGGTVAASADGRFV 586  
QY 541 WAPGDPQPVVAVFGNSWAAASQGVPAQAQIRSDRVNPKTFYALNGNTPYRSTDCGVTF 600  
DB 587 WAPGDPQPVVAVFGNSWAAASQGVPAQAQIRSDRVNPKTFYALNGNTPYRSTDCGVTF 646  
QY 601 QPVAAGLPSGAVGVNMFHVPKGEGLWLAASGLYHSTNGSSWSAITGVSSAVNVGFG 660  
DB 647 QPVAAGLPSGAVGVNMFHVPKGEGLWLAASGLYHSTNGSSWSAITGVSSAVNVGFG 706  
QY 661 KSAPGSSYPAVFVVGITGGVTAIRSDDCGTTWVLINDDQHYGNWQAITGDHANLRV 720  
DB 707 KSAPGSSYPAVFVVGITGGVTAIRSDDCGTTWVLINDDQHYGNWQAITGDHANLRV 766  
QY 721 YIGTNGRGIVYGDIGGAPSG 740  
DB 767 YIGTNGRGIVYGDIGGAPSG 786

RESULT 6  
ADO52313  
ID ADO52313 standard; protein; 957 AA.  
XX AC  
XX ADO52313;  
XX

DT 15-JUL-2004 (first entry)  
XX Acidothermus cellulolyticus avicelase III (Aviii).  
DE Thermostable cellulase; avicelase III; Aviii; cellulose reduction;  
XX biomass degradation; ethanol formation; industrial chemical;  
KW fabric treatment; enzyme.  
KW Acidothermus cellulolyticus.  
OS  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..34  
FT /label= Signal\_peptide  
FT Protein 35..957  
FT /note= "Acidothermus cellulolyticus mature avicelase III"  
FT Domain 35..187  
FT /note = Carbohydrate binding domain type III  
FT Domain 231..870  
FT /note = GH74 catalytic domain  
FT Misc-difference 957  
FT /label= Unknown  
XX  
PN US2004038334-A1.  
XX  
XX 26-FEB-2004.  
XX  
XX 28-JUL-2001; 2001US-00917376.  
XX  
XX 28-JUL-2001; 2001US-00917376.  
XX  
XX (DING/) DING S.  
PA (ADNE/) ADNEY W S.  
PA (VINZ/) VINZANT T B.  
XX (HIMM/) HIMMEL M E.  
XX  
PI Ding S, Adney WS, Vinzant TB, Himmel ME;  
XX  
XX WPI; 2004-203224/19.  
DR N-PSDB; ADO52314.  
XX  
XX Novel thermostable Aviii polypeptide of glycoside hydrolase family and  
PT isolated from Acidothermus cellulolyticus, useful for degrading treated  
PT biomass into simpler forms of carbohydrate.  
XX  
XX Claim 12; SEQ ID NO 1; 19pp; English.  
XX  
XX The invention relates to a thermostable cellulase enzyme, avicelase III  
CC (Aviii) and its nucleic acid sequence. Aviii is useful for reducing  
CC cellulose in a starting material. A thermostable Aviii peptide is useful  
CC for degrading treated biomass into simpler forms of carbohydrate, which  
CC is used in the formation of ethanol or other industrial chemicals. It is  
CC also useful for treating fabrics to remove cellulose-containing stains.  
CC The present sequence is Acidothermus cellulolyticus Aviii protein.

XX SQ Sequence 957 AA;

Query Match 100.0%; Score 4036; DB 8; Length 957;  
Best Local Similarity 100.0%; Pred. No. 1.2e-270;  
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATTQPTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60  
DB 47 ATTQPTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 106  
QY 61 NNMWYGVVSIADPINTNKVAAVGMVYTNWDNDGAILRSSDQATWQITPLPKLGG 120  
DB 107 NNMWYGVVSIADPINTNKVAAVGMVYTNWDNDGAILRSSDQATWQITPLPKLGG 166  
QY 121 NMPGRGMRGLAVDPNNNDILYFGAPSGKGLWRSSTDGATWQMTNPDVGYIANTPTD 180  
DB 167 NMPGRGMRGLAVDPNNNDILYFGAPSGKGLWRSSTDGATWQMTNPDVGYIANTPTD 226  
QY 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNPNFVWSRDGGATWQVPGAPTGF 240



Db 227 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNPNPFWRSRGGATWQAVPGAPTGF 286  
QY 241 IPHKGVPDPVNVHLYIATSNTPGPGVDSGSDVWKFSTSGTWTTRISPPSTDTTANDYGY 300  
Db 287 IPHKGVPDPVNVHLYIATSNTPGPGVDSGSDVWKFSTSGTWTTRISPPSTDTTANDYGY 346  
QY 301 SGLTIDRQHPNTIMVATQISWMPDIIIFRSTDDGGATWTRINDWTSYPNRSRLRYLVDISAE 360  
Db 347 SGLTIDRQHPNTIMVATQISWMPDIIIFRSTDDGGATWTRINDWTSYPNRSRLRYLVDISAE 406  
QY 361 PMLTFGVQPNPVPSPKLGWDEMAIDPFNSDRMLYGTGATLYATNLDLTKWDSGGQIHI 420  
Db 407 PMLTFGVQPNPVPSPKLGWDEMAIDPFNSDRMLYGTGATLYATNLDLTKWDSGGQIHI 466  
QY 421 APWKVGLBEETA VNDLISPPSGAPLISALGDLGGFTHADVAVPSTIIFTSPTFTTGSVDY 480  
Db 467 APWKVGLBEETA VNDLISPPSGAPLISALGDLGGFTHADVAVPSTIIFTSPTFTTGSVDY 526  
QY 481 AELNPSIIVRAGSFPDPSSQPNDRHVAFSTDDGKNWFOGSEPGVTTGGTVAASADGSRFV 540  
Db 527 AELNPSIIVRAGSFPDPSSQPNDRHVAFSTDDGKNWFOGSEPGVTTGGTVAASADGSRFV 586  
QY 541 WAPGDPGPVVYAVGFGNSWAASQGVPANAOIRSDRVNPKTFYALNSGTFFYRSTDDGVTF 600  
Db 587 WAPGDPGPVVYAVGFGNSWAASQGVPANAOIRSDRVNPKTFYALNSGTFFYRSTDDGVTF 646  
QY 601 QPVAAGLPSSGAVGVMFHAPVCKEGDLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 660  
Db 647 QPVAAGLPSSGAVGVMFHAPVCKEGDLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 706  
QY 661 KSAPGSSYPAVFVVGTTGGVTCAYRSDDCGTTWVLINDDQHOYGNWQAITGDHANLRRV 720  
Db 707 KSAPGSSYPAVFVVGTTGGVTCAYRSDDCGTTWVLINDDQHOYGNWQAITGDHANLRRV 766  
QY 721 YIGTNGRGIVYGDIGGAPSG 740  
Db 767 YIGTNGRGIVYGDIGGAPSG 786

RESULT 7  
AD052315  
ID AD052315 standard; protein; 733 AA.  
XX AC AD052315;  
XX DT 15-JUL-2004 (first entry)  
XX DE A. cellulolyticus avicelase III (AviIII) GH4 catalytic domain (CD).  
XX KW Thermostable cellulase; avicelase III; AviIII; cellulose reduction;  
KW biomass degradation; ethanol formation; industrial chemical;  
KW fabric treatment; catalytic domain; CD.  
XX OS Acidothermus cellulolyticus.  
XX PN US2004038334-A1.  
XX PD 26-FEB-2004.  
XX PF 28-JUL-2001; 2001US-00917376.  
XX PR 28-JUL-2001; 2001US-00917376.  
XX PA (DING/) DING S.  
PA (ADNE/) ADNEY W S.  
PA (VINZ/) VINZANT T B.  
PA (HIMM/) HIMMEL M E.  
XX Ding S, Adney WS, Vinzant TB, Himmel ME;  
XX WPI; 2004-203224/19.

PT Novel thermostable AviIII polypeptide of glycoside hydrolase family and  
PT isolated from Acidothermus cellulolyticus, useful for degrading treated  
PT biomass into simpler forms of carbohydrate.  
XX Claim 28; SEQ ID NO 3; 19pp; English.  
XX The invention relates to a thermostable cellulase enzyme, avicelase III  
CC (AviIII) and its nucleic acid sequence. AviIII is useful for reducing  
CC cellulose in a starting material. A thermostable AviIII peptide is useful  
CC for degrading treated biomass into simpler forms of carbohydrate, which  
CC is used in the formation of ethanol or other industrial chemicals. It is  
CC also useful for treating fabrics to remove cellulose-containing stains.  
CC The present sequence is Acidothermus cellulolyticus AviIII protein GH4  
CC catalytic domain (CD).  
XX Sequence 733 AA;  
SQ Query Match 97.1%; Score 3920.5; DB 8; Length 733;  
Best Local Similarity 99.1%; Pred. No. 8.4e-263;  
Matches 733; Conservative 0; Mismatches 0; Indels 7; Gaps 7;  
QY 1 ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVTRDIGMYRWDAAANGRIWILLDWGV 60  
Db 1 ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVTRDIGMYRWDAAANGRIWILLDWGV 60  
QY 61 NNWGVNGVVSIAADPINTNKVWAAVGMVYNSWDNDGAILRSDQCATWQITPLPKLGG 120  
Db 61 NNWGVNGVVSIAADPINTNKVWAAVGMVYNSWDNDGAILRSDQCATWQITPLPKLGG 120  
QY 121 NMPGRGMGERLAVDNNNDNIIYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDT 180  
Db 121 NMPGRGMGERLAV - PNND - ILYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDT 178  
QY 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNPNPFWRSRGGATWQAVPGAPTGF 240  
Db 179 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNPNPFWRSRGGATWQAV - GAPTGF 237  
QY 241 IPHKGVPDPVNVHLYIATSNTPGPGVDSGSDVWKFSTSGTWTTRISPPSTDTTANDYGY 300  
Db 238 IPHKG - FDPVNVHLYIATSNTPGPGVDSGSDVWKFSTSGTWTTRISPV - STDANDYGY 295  
QY 301 SGLTIDRQHPNTIMVATQISWMPDIIIFRSTDDGGATWTRINDWTSYPNRSRLRYLVDISAE 360  
Db 296 SGLTIDRQHPNTIMVATQISWMPDIIIFRSTDDGGATWTRINDWTSYPNRSRLRYLVDISAE 355  
QY 361 PMLTFGVQPNPVPSPKLGWDEMAIDPFNSDRMLYGTGATLYATNLDLTKWDSGGQIHI 420  
Db 356 PMLTFGVQPNPVPSPKLGWDEMAIDPFNSDRMLYGTGATLYATNLDLTKWDSGGQIHI 415  
QY 421 APWKVGLBEETA VNDLISPPSGAPLISALGDLGGFTHADVAVPSTIIFTSPTFTTGSVDY 480  
Db 416 APWKVGLBEETA VNDLISPPSGAPLISALGDLGG - THADVAVPSTIIFTSPTFTTGSVDY 474  
QY 481 AELNPSIIVRAGSFPDPSSQPNDRHVAFSTDDGKNWFOGSEPGVTTGGTVAASADGSRFV 540  
Db 475 AELNPSIIVRAGSFPDPSSQPNDRHVAFSTDDGKNWFOGSEPGVTTGGTVAASADGSRFV 534  
QY 541 WAPGDPGPVVYAVGFGNSWAASQGVPANAOIRSDRVNPKTFYALNSGTFFYRSTDDGVTF 600  
Db 535 WAPGDPGPVVYAVGFGNSWAASQGVPANAOIRSDRVNPKTFYALNSGTFFYRSTDDGVTF 594  
QY 601 QPVAAGLPSSGAVGVMFHAPVCKEGDLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 660  
Db 595 QPVAAGLPSSGAVGVMFHAPVCKEGDLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 654  
QY 661 KSAPGSSYPAVFVVGTTGGVTCAYRSDDCGTTWVLINDDQHOYGNWQAITGDHANLRRV 720  
Db 655 KSAPGSSYPAVFVVGTTGGVTCAYRSDDCGTTWVLINDDQHOYGNWQAITGDHANLRRV 714  
QY 721 YIGTNGRGIVYGDIGGAPSG 740  
Db 715 YIGTNGRGIVYGDIGGAPSG 733











QY	300	YSLGTLDRQHPNTIMVATQISWNPDDTIIFRSTDCGATWTRINWDTSYPNRSLRYVLDDISA	359
Db	314	FGGLGLDLQKPGTILVVASLNSWNPDAQLFRSTDSGTTWSPITAWASPYTETYYYSISTPK	373
QY	360	EPMLTFG-VQPNPPVPS-----PKLGWDEAMADPENSDRMLYGTGTATLYATNDLTKWDS	414
Db	374	APWIKKNFIDVTSPPSDGLIKRLGWNWIESLEIDPTDSNHWLYGTGHTIFGGHDLTNWDT	433
QY	415	GQGIHIAPMWKGLBETAVNDLISPPSGAPLISALDGLGGFTHA---DVTAVPSTIFTSPV	471
Db	434	RHNVSIOQLADGIEEFSVQDLASAPGSGELLAAVGGDNGFTFASRNDLGTSPQTVWATPT	493
QY	472	FTTGTSDVYAEALNPSIIVRAGSFDPSSQPNDRHVAFTDGGKNWFGQSGEPGGVTTGGTVA	531
Db	494	WATSTSDVYAGNSVKSVVVRGNTAGTQQ-----VAISSDGGCATWSIDYAADTSMNGGTTVA	548
QY	532	ASADGSRFVWAPGDPGQPVVYAVGFGNSWAASQGVPAQAQIRSDRVNPKTFYALSNGTFY	591
Db	549	YSADGDTILMSTASSG---VQRSQFGSFASVSLPAGAVIASDKTNSVVFYAGSGSTFY	605
QY	592	RSTDGGVTFQPVAAAGLPSSGAVGVW--FHAVPGKEGDLWLAASGLYHSTNGSSWSAI-	648
Db	606	VSKDTGSSF---TRG-PKLSAGATIRIDIAHPTTAGTLYVSTDVGIFRSTDSGTTFGQVS	661
QY	649	TCGVSSANVNGFGKAPGSSYPAPVFGVTIGGVGTGA--YRSDDCGTTWVLINDDQHQVGNW	706
Db	662	TALTNYQIALGVGS-GSNW-NLYAFGT--GPSGARLYASGDSGASWTDIQSGQGFSGID	717
QY	707	GQAITGDHANLRVYIGTNGRGIVY--GDICGAPSG	740
Db	718	STKVAGSGSTAGQVYVGTNGRGVFAQGTGGGTTGG	753
RESULT 11			
ADH51581			
XX	ADH51581	standard; protein; 818 AA.	
AC	ADH51581;		
DT	25-MAR-2004	(first entry)	
XX	XX	Trichoderma reesei EGVI protein sequence related to ethanol production.	
DE	XX	EGVI protein; endoglucanase; ethanol production; biomass composition;	
KW	XX	fermentation; cellobiohydrolase; detergent production; softening agent;	
KW	XX	cotton fabric; wood pulp degradation; sugar; enzyme.	
XX	OS	Hypocrea jecorina.	
XX	XX	Key Location/Qualifiers	
PH	XX	Misc-difference 1..818	
FT	XX	/label= OTHER	
FT	XX	/note= "OTHER= All Xaa's given in this sequence are	
FT	XX	unknown amino acids, given as J in the specification.	
FT	XX	They should probably be Ile."	
XX	XX	US2003113732-A1.	
PN	XX	19-JUN-2003.	
XX	XX	18-DEC-2001; 2001US-00026994.	
PD	XX	18-DEC-2001; 2001US-00026994.	
XX	XX	(DUNN/) DUNN-COLEMAN N.	
PA	XX	(GOED/) GOEDEGEBUUR F.	
PA	XX	(WARD/) WARD M.	
PA	XX	(YAOJ/) YAO J.	
XX	XX	Dunn-Coleman N, Goedegebuur F, Ward M, Yao J;	
PI	XX	WPI; 2004-106460/11.	
DR	XX	N-PSDB; ADH51580, ADH51583.	
DR	DR		

[illegible]



Db	642	TALNTVQIALGVGS-GSNW-NLYAFGT--GPSGARLYASGDSGASWTDIQSGQFGSID	697	Db	173	S-----VIFDP-----ERNGTIYASATAPQG-MYTHDGGVSWEPVAGQPSWLNRT	218
Qy	707	GOAITGHDANLRRVYIGTNGRGIYV--CDIGGAPSG	740	Qy	242	-----PHKGVDPVNVHLYIATISNTGGPGVDSGSSGVWKFVSFVTSGTWTRI-	285
Db	698	STKVAGSGTAGQVYVGTNGRGRVYAGTQVGGGTGG	733	Db	219	TGAPDKKPSIAPQPMKVALTP--NFLYVIYADYPGPGVTFGEVWQNRISGAWDDIT	276
RESULT 12				Qy	286	-----SPVSTDTANDYFGYSGLTIDROHPNTIMVATOISWMPDTI---IFRSTDDGGAT	336
ADH19120				Db	277	PRVGNSSPAPYNNQTFPAGGFCGLSVDATNPRLVIT-LDRDPGAPALDSIYLSLTDAGAT	335
XX	ID	ADH19120 standard; protein; 789 AA.		Qy	337	WTRIWDWTS-----YPNRSURYVLDISAPMLTFQVQP-----NPPVPSP---	379
XX	AC	ADH19120;		Db	336	WKDVTQLSSPSNLEGNWGHPTNAARY-KDGTVPVPLWDFNNGPQWGGYGAPHGTPGLTKFG	394
XX	DT	11-MAR-2004 (first entry)		Qy	380	WMDEMAIDPNSDRMLYGTGATLYATNDLTK-----WDSGGQIHLAP-----MVKGLEETA	431
XX	DE	Geotrichum sp M128 xyloglucan oligosacch-degrading enzyme mat protein.		Db	395	WWMGAVLIDPFPNPHLMYGTGATIWATDLSRVEKDW-----APSWYLQIDGIEENA	446
XX	KW	xyloglucan oligosaccharide degradation; enzyme; genetic engineering;		Qy	432	VNDLISPPSGAPLISALGDLGFTHADVTAVPSTIFTSPVFTTGTSDVYAEINPSIIVRA	491
XX	KW	mature.		Db	447	ILSLRSPKSGAALLSGIGDISGMKHDLLTK-PQKMFAGPQFNSLDSIDAAGNFPNVVRA	505
XX	OS	Geotrichum sp. M128.		Qy	492	GSPDPSSQPNDRHVAFSTDDGKNW--FQSGEPGGVTT---GGTVAASADGSRFPWAP--G	544
XX	FN	EP130844-A2.		Db	506	GSSGHEYDSACARGAYATDGGDAWTFPTCPPEGMNASHYQGSTIAVDASGSQIVWSTKLD	565
XX	PD	08-OCT-2003.		Qy	545	DPGQPVVYAVGFGNSWAASQGVPA-----NAQIRSDRVNPKTFVALSNGTFRVSTDDGV	598
XX	PF	25-MAR-2003; 2003EP-00251866.		Db	566	EQASGPWYSHDYGKTWS-----VPAGDLKAQTANVLSKVDQDTFYATDGGKFFVSTDDGK	621
XX	PR	25-MAR-2002; 2002JJP-00083433.		Qy	599	TFQPVAGLPSGAGVGMFHAVPGKEGDLWL-AASSGLYHSTNGSSWSAI-TGVSSAVN	656
XX	PA	(NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.		Db	622	SYAAKAGLVN--GTSMLPAVNPVWAGDVWVVPESGLFHSSTDFGASFTRVGTANATLVS	679
XX	PI	Yaoi K, Mitsuishi Y;		Qy	657	VGFGKS-----APGSSYPVAVVVGTT--IGVTVGAYRSDDCGTTWVLINDDQHOYGNWQAI	710
XX	DR	WPI; 2004-100948/11.		Db	680	VGAPKSKDGKASAPSAVFIWGTDPKGSIDIGLYRSDDNGSTWTRVNDQEHYSG-PTMI	738
XX	PT	Novel xyloglucan oligosaccharide-degrading enzyme with a different		Qy	711	TGDHANLRVVYIGTNGRGIYVYDI	734
XX	PT	degradation mechanism from known enzymes.		Db	739	EADPKVYGRVYLTNGRGIYVADL	762
XX	PS	Claim 2; SEQ ID NO 14; 39pp; English.					
XX	CC	The invention relates to a novel xyloglucan oligosaccharide-degrading		RESULT 13			
XX	CC	enzyme with a different degradation mechanism from known enzymes. The		ADH19118			
XX	CC	primary structure and polynucleotide structure of the xyloglucan		ID	ADH19118 standard; protein; 812 AA.		
XX	CC	oligosaccharide-degrading enzyme provided by the present invention allows		XX	ADH19118;		
XX	CC	high-purity polypeptides having a xyloglucan oligosaccharide-degradation		XX	11-MAR-2004 (first entry)		
XX	CC	activity to be prepared at a low cost through a genetic engineering		DT	Geotrichum sp M128 xyloglucan oligosacch-degrading enzyme pre protein.		
XX	CC	process. The current sequence is that of the Geotrichum sp. M128		DE	xyloglucan oligosaccharide degradation; enzyme; genetic engineering;		
XX	CC	xyloglucan oligosaccharide-degrading enzyme mature protein of the		KW	precursor.		
XX	CC	invention.		XX	Geotrichum sp. M128.		
XX	SQ	Sequence 789 AA;		XX	EP130844-A2.		
Query Match		27.7%; Score 1117; DB 8; Length 789;		XX	08-OCT-2003.		
Best Local Similarity		35.0%; Pred. NO. 1.7e-68;		XX	25-MAR-2003; 2003EP-00251866.		
Matches 281; Conservative 124; Mismatches 279; Indels 120; Gaps 31;				XX	25-MAR-2002; 2002JJP-00083433.		
Qy	6	YTNWNAIIGGGGVGDGIVFNEGAPGLYVRVTDIGMYRWDAAANGRWIPLLDWGVNNGWY	65	XX	(NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.		
Db	4	YEFKRVAIIGGGGIYTGIVAHFKDLYARDIGAYRWADAGTSKWIPNDIFIAQDNWI	63	XX	Yaoi K, Mitsuishi Y;		
Qy	66	NGVYSIADPINTNKNVAAVGMYT-NSWDPNDGAILRSSDQAGATWQITPLFPKLGNNMPG	124	XX	WPI; 2004-100948/11.		
Db	64	MGTESIALDPNNPDLRYLAQRYGVDEW---AAPVYSEDGRQSFTIYESFPNGANDWG	119	XX	N-PSDB; ADH19119.		
Qy	125	RGMERLAVDPNNNILYFGAPSGKGLWRSTDSGATWSQMTNFPDPVGYIANPTDTTGYQ	184	XX	Novel xyloglucan oligosaccharide-degrading enzyme with a different		
Db	120	RNNGERLAVNPNFNSNEVWMTGRT-EGIWKSDRAKTWNTVTSIPDAF-----TNGIGYT	172	XX	degradation mechanism from known enzymes.		
Qy	185	SDIOGVVWVAEDKSSSSISQASKTIFVGVADPNPNFVWSRDGATWQAVPGATGFI---	241	XX	Claim 2; SEQ ID NO 14; 39pp; English.		



PT Novel xyloglucan oligosaccharide-degrading enzyme with a different  
 PT degradation mechanism from known enzymes.  
 XX Claim 5; SEQ ID NO 12; 39pp; English.  
 XX

The invention relates to a novel xyloglucan oligosaccharide-degrading  
 CC enzyme with a different degradation mechanism from known enzymes. The  
 CC primary structure and polynucleotide structure of the xyloglucan  
 CC oligosaccharide-degrading enzyme provided by the present invention allows  
 CC high-purity polypeptides having a xyloglucan oligosaccharide-degradation  
 CC activity to be prepared at a low cost through a genetic engineering  
 CC process. The current sequence is that of the Geotrichum sp. M128  
 CC xyloglucan oligosaccharide-degrading enzyme precursor protein of the  
 CC invention.

XX Sequence 812 AA;  
 SQ

Query Match 27.7%; Score 1117; DB 8; Length 812;  
 Best Local Similarity 35.0%; Pred. No. 1.7e-68;  
 Matches 281; Conservative 124; Mismatches 279; Indels 120; Gaps 31;

QY 6 YTSNVAIGGGFVDGIVFNEGAPGILYVRTDYGMYRWDAAANGRWIPLLDWVGNNWGY 65  
 DB 27 YEFKVAIGGGYITGIVAHPTKDLLYARTDIGGAYRWDACTSKWPLNDFIEAQDMNI 86  
 QY 66 NGVVSIAADPINTNKVAAVGMYT-NSWDPNDGAILRSSDQATWQITPLPKLGGNMPG 124  
 DB 87 MGTESIALDPNNPDRLLYLAQGRVGDW----AAFVSEDRGQSFTIYESPPFMGANDMG 142  
 QY 125 RGMGERLAVDPNNDNLYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDTTGYQ 184  
 DB 143 RNNGERLAVPNFNSNEVMWGTRT-EGITWSSDRAKTWNVTISPDAF-----TNGIGYT 195  
 QY 185 SDIQGVVWVAFDKSSSLGQASKTIFVGVADPNPNFWSRDGGATWQAVPGAPTGF1--- 241  
 DB 196 S-----VIFDP-----ERNGTIYASATAPOG-MYVTHDGGVSWEPVAGQPSWLNRT 241  
 QY 242 -----PHKGVDPVNHVLIYATNTGGPYDSSGSDGVWKFPSVTSGTWTRI- 285  
 DB 242 TGAFPDKKPAIAPQPMKVALTP--NFLYVTYADYPGPGVTFGEVWRQNTSGAWDDIT 299  
 QY 286 -----SPVSTDTANDYFGYSLGILDRQHNPTIMVATQISWPDIT---IFRSTDDGAT 336  
 DB 300 PRVGNSSPAPYNNQTFPAGGFCGLSDVATNPNRLVIT-LDRDGPALDSIYLSYDAGAT 358  
 QY 337 WTRIDWTS-----YPNRSRLYLDISAEPWLTFCGVQ----NPPVPS----KLG 379  
 DB 359 WKDVTQLSPSNLEGNWGHPTWAARY-KDGTVPWLDNFNNGQWGGYGAPHGTFGLTKFG 417  
 QY 380 WMDEAMADPPFNSDRMLYGTGATLYATNDLTK----WDSGGQIHIA----MVKGLEETA 431  
 DB 418 WMSAVLIDPPFPEHLWMTGTGATIWATDLSRVEKDW-----APSWYLQIDGIEENA 469  
 QY 432 VNDLISPPSGAPLISALGDLGFTHADVTAVPSTITFTSPVTTGTSVDYAEINPSIIVRA 491  
 DB 470 ILSLRSPKSGAALLSIGDIGSKHDDLTK-PQKMGAPQFNSLSDISAAGNFPNPNVRA 528  
 QY 492 GSFPDSSQPNDRHRVAFSTDDGKNW--FQSGEPGGVTT---GCTVAASADGSRFWAP--G 544  
 DB 529 GSGHEGYDSACARGAYATDGGDAWITFTCTPPGMASHYQGGTIIADVASGSQIVWSTKLD 588  
 QY 545 DPQOPVYAVVAVFGNSWAASQGVPA-----NAQIRSDRVNPKTFYVSLNSGTFRSTDGCV 598  
 DB 589 EQASGPWYSHDYGTWS-----VPAGDLKAQTANVLSDKVDQGTFFATDGGKFFVSTDDGK 644  
 QY 599 TFOVVAAGLPSSGAVGVMFHVPKREGDLWL--AASGLYHSTNGSSWSAI--TGVSASV 656  
 DB 645 SYAAGKAGLVT--GTSLMPAVNPWVAGDVWVPVPEGGLFHSDFHSGFTRVGTANATLVS 702  
 QY 657 VGFEGS-----APGSSYPVAVVGT--IGGVTCAYSDDCGTTWLVINDDOHOGNMGQAI 710  
 DB 703 VGAPKSKDGKKASAPSAVFTWGTDPKPSDILGLYRSDDDNGSTWTRVNDQEHYSG--PTWI 761

QY 711 TGDHANLRVYTGTRNGRIVYGD1 734  
 DB 762 EADPKVYGRVYLTGTRNGRIVADL 785

RESULT 14  
 ADH19124  
 ID ADH19124 standard; protein; 826 AA.  
 XX  
 AC ADH19124;  
 XX  
 DT 11-MAR-2004 (first entry)  
 XX  
 DE Geotrichum sp xyloglucan oligosacch-degrading enzyme-derived protein.  
 XX  
 KW xyloglucan oligosaccharide degradation; enzyme; genetic engineering;  
 KW His-tag.  
 XX  
 OS Synthetic.  
 OS Geotrichum sp. M128.  
 XX  
 PN EP1350844-A2.  
 XX  
 PD 08-OCT-2003.  
 XX  
 PF 25-MAR-2003; 2003EP-00251866.  
 XX  
 PR 25-MAR-2002; 2002JP-00083433.  
 XX  
 PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.  
 XX  
 PI Yaoi K, Mitsuishi Y;  
 XX  
 DR WPI; 2004-100948/11.  
 DR N-PSDB; ADH19123.  
 XX  
 PT Novel xyloglucan oligosaccharide-degrading enzyme with a different  
 PT degradation mechanism from known enzymes.  
 XX  
 PS Claim 6; SEQ ID NO 18; 39pp; English.  
 XX

The invention relates to a novel xyloglucan oligosaccharide-degrading  
 CC enzyme with a different degradation mechanism from known enzymes. The  
 CC primary structure and polynucleotide structure of the xyloglucan  
 CC oligosaccharide-degrading enzyme provided by the present invention allows  
 CC high-purity polypeptides having a xyloglucan oligosaccharide-degradation  
 CC activity to be prepared at a low cost through a genetic engineering  
 CC process. The current sequence is that of the Geotrichum sp. M128  
 CC xyloglucan oligosaccharide-degrading enzyme-derived protein of the  
 CC invention which comprises the wild-type enzyme sequence plus a C-terminal  
 CC region including a Histidine tag originating from vector pET29a(+).

XX Sequence 826 AA;  
 SQ

Query Match 27.6%; Score 1114; DB 8; Length 826;  
 Best Local Similarity 35.0%; Pred. No. 2.8e-68;  
 Matches 281; Conservative 123; Mismatches 280; Indels 120; Gaps 31;

QY 6 YTSNVAIGGGFVDGIVFNEGAPGILYVRTDYGMYRWDAAANGRWIPLLDWVGNNWGY 65  
 DB 5 YEFKVAIGGGYITGIVAHPTKDLLYARTDIGGAYRWDACTSKWPLNDFIEAQDMNI 64  
 QY 66 NGVVSIAADPINTNKVAAVGMYT-NSWDPNDGAILRSSDQATWQITPLPKLGGNMPG 124  
 DB 65 MGTESIALDPNNPDRLLYLAQGRVGDW----AAFVSEDRGQSFTIYESPPFMGANDMG 120  
 QY 125 RGMGERLAVDPNNDNLYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDTTGYQ 184  
 DB 121 RNNGERLAVPNFNSNEVMWGTRT-EGITWSSDRAKTWNVTISPDAF-----TNGIGYT 173  
 QY 185 SDIQGVVWVAFDKSSSLGQASKTIFVGVADPNPNFWSRDGGATWQAVPGAPTGF1--- 241  
 DB 174 S-----VIFDP-----ERNGTIYASATAPOG-MYVTHDGGVSWEPVAGQPSWLNRT 219



QY 242 -----PHKGVDPVNVHLYIATSNITGGSPYDSSGDWKFVSITSGTWRI- 285  
 DB 220 TGAFPPKKPASIAPQPKVALTP--NFLYVYADYPGVGWTFKVRQNRISGAWDDIT 277  
 QY 286 -----SVPSTDTANDYFGSLGTLIDRQHPNTIMVATQISWPDPTI---IFRSTDGAT 336  
 DB 278 PRVGNSSPAPNNQTFPAGGFCGLSDATNPNRLVWIT-LDRDFGPALDSIYLSYLDAGAT 336  
 QY 337 WTRLDWTS-----YPNRSLRYLVDISAEPWLTFGVOP-----NPPVPSP---KLJ 379  
 DB 337 WKDTQLSSPNLEGNWGHPTNAARY-KDGTVPWLDFFNPGQGGYCAPHGTFGLTKFG 395  
 QY 380 WMDERMAIDPNSDRMLYGTGATLYATNDLTK-----WDSGGQIHIAP-----MVKGLEETA 431  
 DB 396 WMSAVLIDPNEHMYGTGATIWAIDTLRVEKDW-----APSWLIQIDGIBENA 447  
 QY 432 VNDLISPPSGAPLISALGDLGGFTHADVAVPSTIFTSPTVTGTSVDYAEINPSIIVRA 491  
 DB 448 ILSLRSPKSGAALLSGIGDISGMKHDDLTG-PQRMFGAPQFSLNDSIDAAGNFFNVPVRA 506  
 QY 492 GSFDPSSQPNDRHVAFTSDGKNW--FQGSBPGGVTT--GGTVAASADGSRFVWAP--G 544  
 DB 507 GSSGHEYDSACARGATDGGDAMTIFPTCPGMNASHYQGSTIAVDASGSIQVWSTKLD 566  
 QY 545 DPGQPVVYAVGFGNSWAASQGVPA-----NAQIRSDRVNPKTFYALSNGTFYRSTDCGV 598  
 DB 567 EQASGPWYSHDYGKTS-----VPAGDLKAQATNVLSDKVQDGTFTVATDGGKFFVSTDGK 622  
 QY 599 TFQPVAAGLSSGAVGVNMFHAPVQKBDLWL-AASSGLYHSTNGSSWSAI-TGVSSAVN 656  
 DB 623 SYAAKAGGLVT--GTSLMPAVNPVWAGDVVVPVEGGLFHSHTDFGASPTRVGTANATLVS 680  
 QY 657 VGFGKS-----APGSSYPVAVFVGT--IGVVTGAYRSDDCGTTWVLINDDOHOYGNWGOAI 710  
 DB 681 VGAPKSKSDGKKASAPSAVFVWGTGDKPGSDIGLYRSDDDSGSTWTRVNDQEHNSG-PTMI 739  
 QY 711 TGDHANLRRVYIGTNGRIVYGD I 734  
 DB 740 EADPKYGRVILGTNGRIVYADL 763

RESULT 15  
 ID ADR90298 standard; protein; 756 AA.  
 XX AC ADR90298;  
 XX DT 16-DEC-2004 (first entry)  
 XX DE Geotrichum sp. M128 xyloglucanendohydrolase mature protein.  
 XX KW endo-type xyloglucanendohydrolase;  
 KW xyloglucan-specific endo-b-1,4-glucanase; glucose; xylose; xyloglucan;  
 KW plant cell differentiation; mature; enzyme.  
 XX OS Geotrichum sp. M128.  
 XX PN JP2004261037-A.  
 XX PD 24-SEP-2004.  
 XX PF 28-FEB-2003; 2003JP-00053286.  
 XX PR 28-FEB-2003; 2003JP-00053286.  
 XX PA (DOKU-) DOKURITSU GYOSEI HOJIN SANGYO GIJUTSU SO.  
 XX DR WPI; 2004-665466/65.  
 XX DR N-PSDB; ADR90297.  
 PT Novel xyloglucan hydrolase derived from Geotrichum genus, having ability to hydrolyze beta-1, 4-glycoside linkage of glucose residue in

PT xyloglucan, useful for elucidating structure or function of xyloglucan.  
 XX Claim 2; SEQ ID NO 15; 86pp; Japanese.  
 CC The invention relates to a novel xyloglucanendohydrolase (xyloglucan-specific endo-b-1,4-glucanase) enzyme derived from the Geotrichum genus and having the ability to hydrolyse the beta-1,4-glycosidic linkage of a glucose residue, but not that of a xylose residue contained in a xyloglucan. The polypeptide of the invention may be useful for elucidating the structure or function of a xyloglucan having an important role in the differentiation of a plant cell. The current sequence is that of the Geotrichum sp. M128 xyloglucanendohydrolase mature protein of the invention.  
 XX Sequence 756 AA;  
 SQ  
 Query Match 24.3%; Score 979; DB 8; Length 756;  
 Best Local Similarity 33.1%; Pred. No. 5.7e-59;  
 Matches 264; Conservative 122; Mismatches 279; Indels 132; Gaps 33;  
 QY 11 VAIGGGFVDGIVFNEGAPGILYVTRDIGMYRMDAANGRWIPLLDWV-----GWNWGY 65  
 DB 8 VTISGGFISGLVAHPTEKDLIYARTDIGTYRNAAKWEWEPITDFIINNALAGNANL 67  
 QY 66 NGWVSIADPTNTNKKVAAVGMVNTNSWDPNDGAILRSSDQGWATQITPLPFKLGMPGR 125  
 DB 68 LGTESIALDPNDRLYLAQGDYV-QWDP-WAAFLVSDDRGKTFKQYRSPVPMGANDMGR 125  
 QY 126 GMBERLAVDPNNDILLYFAGSGKGLWRSTDSGATWSQMTNPPDPVGTIANTDTTGYQS 185  
 DB 126 NGERLAVNPHTWDELWFGSRT-QGLWRSTDRQAQWNRNQLPDSSTV-----GI-- 174  
 QY 186 DIQGVVWVAFKSSSLGQASKTIFVGVADPNPNVFWSRDGGATQWAVPGAPT----- 238  
 DB 175 ---GLISVIFDPKVGVTAYASHAVGGLW-----VTW--DGGANNSVGGQTFQNSDWTK 224  
 QY 239 -----GFIPHKGVDPVNVHLYIATSNITGGSPYDSSGDWKFVSITSGTWRI 285  
 DB 225 SIVAASGTAIQSSGFLPIKIALGK-NGRLYITYSDAPGFWGVLYGEVMSYDPTNGNWKHI 283  
 QY 286 SPVSTDTANDY-----FGYSLTIDRQHPNTIMVATQISWPDPTIIFRSTDCG 334  
 DB 284 T--PSREGANTYPAPTGNKKVVPVGWNGISVG--NGDTVVVSTLDANGEDS-VYLSRDAG 338  
 QY 335 ATWTRIMDWTSPNRSRLRYLVDISA-----PMLTFGVQVNPVPPSPKLGWDEAMAI 387  
 DB 339 NSWDKLGKLTTPAGAGNSQKESDAKLRNGTLPWLSFQNRGSGIV---GFGWLAAILL 395  
 QY 388 DPFNSDRMLYGTGATLYATNDLTKWDSGGQIHIAPM-----VKGLEETAVNDLISPPSG-A 442  
 DB 396 DPF-SDRLLYGTGAVIATDAVSRADS-----NQAFSWYINTEGIEETAILVLKSPAGPA 450  
 QY 443 PLISALGDLGGFTHADVAVPSTIFTSPTVTGTSVDYAEINPSIIVRAGSFD--PSSQP 500  
 DB 451 HLFSGMYDLGMRHDDDF-SVPQPMYSKPTFSSTDLGLDFAGRAANLARYGRNDHPDAGVA 509  
 QY 501 NDRHVAFTSDGKNW--FQGSBPG-GVTTGTGTVAAASADGSRFVWAPGD-PGQPVVYAVGF 556  
 DB 510 GCTQGAYTNSGDSWTLFTQCVPSLEVNGNGTIVAGADKTFVWSFSAKDGKPYTSSDY 569  
 QY 557 GNSWAASQGVPANAO-IRSDRVNPKTFYALSNGTIFYRSTDGVTTFQPVAAGLPSSGAVGV 615  
 DB 570 KGTWTAPSGLSKQTTGIAADRQVANTFYVYVEGDFVSTDCGSKSYTKKGNGLPCCWT--- 626  
 QY 616 MFHAVPG-----KEGDLWLAASS-GLYHSTNGSSWSAITGVSSAVNVPVPMGKSAAGSSYPA 670  
 DB 627 -YTGTPVTNLRAGELWVSVKGVYHSTDFGNTFTALAGSSSLN-----PA 673  
 QY 671 VFWVGT-----IGGV-----TGAYRSDDCGTTWVLINDDOHOYGNWGOAITGD 713  
 DB 674 VFSIGAPQTPNATETFLWGLIPSAQSQPEGLYMSDNGGLWTRLNDDAHNYGG-ATVISGD 732  
 QY 714 HANLRRVYIGTNGRIV 730



Db 733 PRIYGRVYIGNMGRGII 749  
||||| |||||

Search completed: October 5, 2005, 08:03:10  
Job time : 145.761 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model  
Run on: October 5, 2005, 07:48:13 ; Search time 20,5336 Seconds  
(without alignments)  
3467.504 Million cell updates/sec

Title: US-09-917-376-3  
Perfect score: 4036  
Sequence: 1 ATTQPYTWSNVAIGGGFVD.....YGTNGRGIVYGDIGGAPSG 740

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues  
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 79:\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2371	58.7	890	2 T35237	probable secreted
2	2009	49.8	839	2 D97013	probably secreted
3	1680	41.6	856	2 T00349	Avicelase III - As
4	1160.5	28.8	707	2 F72393	hypothetical prote
5	238	5.9	2468	2 A83412	hypothetical prote
6	202.5	5.0	2174	2 E95965	hypothetical glyci
7	189	4.7	3972	2 S75251	hypothetical prote
8	186.5	4.6	4199	2 S76412	hypothetical prote
9	182	4.5	1904	2 T13256	tail-host specific
10	175.5	4.3	2523	2 F70846	probable PPE prote
11	172	4.3	993	2 AE1905	outer membrane sec
12	171	4.2	908	2 AE2254	hypothetical prote
13	170.5	4.2	5188	2 B85547	probable RTX famil
14	168.5	4.2	5291	2 F90696	hypothetical prote
15	168	4.2	2124	2 A28452	proteoglycan core
16	167.5	4.2	2554	2 AB3528	extracellular seri
17	166.5	4.1	902	2 H87323	hypothetical prote
18	166.5	4.1	980	2 H90681	probable flagellin
19	166	4.1	699	2 D70533	hypothetical prote
20	166	4.1	1468	2 A44345	nucleoporin - rat
21	165.5	4.1	980	2 D85532	probable structura
22	163.5	4.1	618	2 T49741	related to stress
23	163.5	4.1	13055	2 T16580	hypothetical prote
24	163	4.0	1684	2 S10789	amylase A-180 - al
25	160.5	4.0	3716	2 E70969	probable PPE prote
26	160	4.0	1341	2 H98323	hypothetical prote
27	160	4.0	3624	2 AD0835	large repetitive p
28	158.5	3.9	1441	2 B86807	hypothetical prote
29	158	3.9	2232	2 T34434	hypothetical prote

## ALIGNMENTS

### RESULT 1

T35237  
Probable secreted cellulase - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 09-Jul-2004  
C:Accession: T35237  
R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, September 1998  
A:Reference number: Z21572  
A:Accession: T35237  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-890 <SEE>  
A:Cross-references: UNIPROT:O86727; EMBL:AL031515; PIDN:CAA20642.1; GSPDB:GN00070; SCOE  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOE:SC5C7.30C

Query Match 58.7%; Score 2371; DB 2; Length 890;  
Best Local Similarity 57.7%; Pred. No. 5.9e-131;  
Matches 431; Conservative 109; Mismatches 183; Indels 24; Gaps 10;

QY	6	YTSNVAIGGGFVDGIVNEGAPGILYVTRTDIGGMYRWDAANGRWIPLLDVGNWNGY	65
DB	41	YTKNARIIDGGFVPGIVFNRTKDLAYARTDIGGAYRWOEEHTWPLLDHVGWDWGH	100
QY	66	NGVVSIAADPINTNKYAAVGMVYNSWDPNDDGAILRSSDQATWQITPLPFLGNNPGR	125
DB	101	TGVVALASDAVDPRVYAAVGTITNDWDPNGAVLSADRGSWEKADLPFLGNNPGR	160
QY	126	GMGERLAVDPNNNIIYFGAPSGKGLWRSTDGATWSQMTNFFDVGVYIANPTTTGVQS	185
DB	161	GMGERLAVDPHDNDVLYLGAPSGHGLWRSTDAGVTWSEVTAFFPNPGYAOQNDTSGVAS	220
QY	186	DIQGVVAVAPDKSS-SSLQOAKSTIFVGVADPNPVPWSDGGATWQAVPCAPGFPKH	244
DB	221	DNQGITVVTPESTGGAGTATRTLYVGVADKENAVYRSTDAGATWERLAGOQTYLAHK	280
QY	245	GVFDPVNHVLYATSNTPGVDGSGGVKFSVTSGTWTRISVPVSTDTANDYFGYSGLT	304
DB	281	GVLDAGNXYLYLAYSDTGGYDGGKLYRIATATGTWDTISPAAEADT---YGFSGLT	337
QY	305	IDRQHPNTIMVATQISWPDITIIFRSTDGGATWTRIMDWTSYPNRSRYVLDISAEPMLT	364
DB	338	VDRQRCTVWATYSSWPDITQIFRSTDGATWSQAWSYTSYPDRENRYTMDVSSPMLT	397
QY	365	FGVQPNPVPSPKLGWMDAEMADPNFSDMLYGTGATLYATNDLTW-DSSGGIHIAPM	423
DB	398	WGANPAPPEQTPKLGWMTALEIDPFDSDRMYGTGATVGTENTLTWMDDEGGTFAVEPM	457
QY	424	VKGLEETAVNDLISPPSGAPLISNADLGCGFTHADVTAVPSTIFTSVPVFTGTSVDVAEL	483



Db 458 VRGLBETAVNDLASPPSGAPLLSALGDVGGFPHRTSLTEVPSSMMYSPNFTSTTSJLDFAE 517  
QY 484 NPSIIVRAGSFPPSQPNDRHVAFTDGGKWNFGQSEPGVTTGTVAAASADGSRFVWAP 543  
Db 518 KDVVVRAGNLD--SGP---HIAFTDNGANWFGTDPGSGGGTVAAGADGSRFVWSP 572  
QY 544 GDPQPVVYAVFGNSWAAASQGVPAQAQIRSDRVNPKTFYALNSGTFYRSTDGGVTPQPV 603  
Db 573 --EGAGVQVTTGFTGTSWQASTGLPAGAI VESDRVNPATPYGKGRFYVSTDGGATFTAS 630  
QY 604 AA-GLPSSGAVGMFHVPGKEDLWLAASS-----GLYHSTNGSSWSAITGVSSAVNV 657  
Db 631 AATGLPAGD--GVRKALPFGGEGDVLGAGAADGYGLWHSTDGGTFTRLPGVDAADTV 688  
QY 658 GFKGAPGSSYPVAVFVGTGTCAYRSDDCGTTVWLINDDOHOYGNWQAITGDHANL 717  
Db 689 GFKAAPGASQYTLTSAIEIGVGRGIFRSTDAGATWTRVNDDAHQWGTGAITGDPRVY 748  
QY 718 RRVYGTNGRGIVYGD1-----GGAPSG 740  
Db 749 GRVYVATNGRGIVYGDTSDTGGGTDPG 775

RESULT 2  
D97013  
probably secreted sialidase, several ASP-boxes and dockerin domain [imported] - Clostrid  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 09-Jul-2004  
C:Accession: D97013  
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
J.; Dally, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: D97013  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-839 <CUR>  
A:Cross-references: UNIPROT:Q97KK0; GB:AE001437; PIDN:AAK78895.1; PID:g15023820; GSPDB:G  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC0919

Query Match 49.8%; Score 2009; DB 2; Length 839;  
Best Local Similarity 50.9%; Pred. No. 7.8e-110; Mismatches 240; Indels 16; Gaps 12;  
Matches 377; Conservative 107; Mismatches 240; Indels 16; Gaps 12;

QY 3 TQPYTWSNVAIGGGFVDGIVNEGAPGILYVRTDGGMYRWDAAANGRWIPLLDWVGNW 62  
Db 37 SQGYKWDNAKIGAGGVVPAVIFNKTEKDLIYARTDMGAYRWDXKANNKWIPTD--GFS 94  
QY 63 WGYNGVSTAADPINTNKAAGVMTNSWDPNDAAILRSSDQATWQITPLPFLKGNM 122  
Db 95 WMLGCESTADPIDTNRVYIAAGLYTNDQDENAYILSSODKGNTRKYQLPFFKVGNN 154  
QY 123 PGRGGERLAVDPNDNIIYFCAPSGKGLWRSTDSGATWSQMTNPDVGTYIANTDTTG 182  
Db 155 PGRNGERLQIDPNDKILYLARGSGNLWKSSEDYGTWKSVDNFPDGTGYQDQFNE-- 212  
QY 183 YQSDIQGVVVAADFSSSLGQASKTIFVGVAD--PNNPVFWRDGGATWQAVPGAPTGI 241  
Db 213 YTADKGVVWVEITDPSGTGKSPQTQVMYGAADKGNIIYVINDGGKWSA VKGQPKGYL 272  
QY 242 PHKGVPDPNVHLYIATNSGTPYDGSQGVVWKFVSQVTSQWTRISPVSTDTANDYFGVS 301  
Db 273 PHHGLI-ASDGLMYISYNTCPYDGSQGVVWKFVSQVTSQWTRISPVSTDTANDYFGVS 328  
QY 302 GLTIDRQHPNTIMVATQISWPDITIIFRSTDCGATWTRIMWTSYVNSLRVLDISAEP 361  
Db 329 GISVDAQNPNNVAVATLNRWPDDEIYRSTDAKGTWKPIMDWNGVYNTLGYNLDSAQ 388  
QY 362 WLTFFG-VQNPVPSPKLCWMDENMAIDPFNSDRMLYGTGATLYATNLDLTKWDSGQHI 420

Db 389 WLDWKTGVTGTPPDLIVKLGMMDLEIDPFNSDRMFPYGTGATLYGTDDLTNNWKGKKNVDI 448  
QY 421 APMVKGLBETAVNDLISPPSGAPLLSALGDGLGGFTHADVTAVPSTFTSFVFTTGSVDY 480  
Db 449 SVKANGIEECANDVNVVFTKGAQLLSAVGDDCGFFYHDDITTKVPSKMMTTTFNFSATSIDY 508  
QY 481 AELNPSIIVRAGSFPPSQPNDRHVAFTDGGKWNFGQSEPGVTTGTVAAASADGSRF 539  
Db 509 AESVFNFRVVRGVNVDTSKNQDKCGISYDGGKWNFSAGNSISGVYKAGTVAAGADAKTI 568  
QY 540 WMAPGDPQCPVVYAVFGNSWAAASQGVPAQAQIRSDRVNPKTFYALNSGTFYRSTDGGVT 599  
Db 569 WNSP--EEGANAAYSTDNKNGKTPCGLPQGAQVAKVSDRVNPKTFYGLNKFYISTDAGAT 627  
QY 600 F-QPVAAGLPSGAGVGMFHVPGKEDLWLA--ASSGLYHSTNGSSWSAITGVSSAVNV 657  
Db 628 FTQSSQTLGTPKKG-GI-FKTVIGHGDIWIAGKDKLWHSTDSGATFTKVGSDVADSTV 685  
QY 658 GFKGAPGSSYPVAVFVGTGTCAYRSDDCGTTVWLINDDOHOYGNWQAITGDHANL 717  
Db 686 GLGKSKTDDGYPAIYMDATIDGTAGIFRSDDDEGATWTRVNDDAHQYGSPTYCITGDPNKY 745  
QY 718 RRVYGTNGRGIVYGD1GGA 737  
Db 746 GRVYVATNGRGIVYGDIDGS 765

RESULT 3  
T00349  
Avicelase III - Aspergillus aculeatus  
C:Species: Aspergillus aculeatus  
C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 09-Jul-2004  
C:Accession: T00349  
R:Arai, M.; Takada, G.; Kawaguchi, T.; Sumitani, J.  
submitted to the EMBL Data Library, June 1998  
A:Description: Avicelase III from Aspergillus aculeatus.  
A:Reference number: Z14141  
A:Accession: T00349  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-856 <ARA>  
A:Cross-references: UNIPROT:O74170; EMBL:AB015511; NID:dl199887; PID:dl029971  
C:Genetics:  
F:823-854/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 41.6%; Score 1680; DB 2; Length 856;  
Best Local Similarity 46.2%; Pred. No. 1.3e-90;  
Matches 346; Conservative 113; Mismatches 258; Indels 32; Gaps 17;

QY 1 ATTQPYTWSNVAI--GGGGFVDGIVNEGAPGILYVRTDGGMYRWDAAANGRWIPLLDWVG 59  
Db 21 AASQAYTWKVVVGGGGFTPGIVFNPVSAKVAYARTDGGAYRLN--SDTWTPLMDWVG 79  
QY 60 ---WNNWYGVGVVSTAADPINTNKAAGVMTNSWDPNDAAILRSSDQATWQITPLP 116  
Db 80 NDTWHDW---GIDALATDPVDTRVYVAVGMYTNEWDPNVGSILRSTDQGDITWETKLP 136  
QY 117 KLGNMPPGRGGERLAVDPNDNIIYFCAPSGKGLWRSTDSGATWSQMTNPDVGYIYAN 176  
Db 137 KYGVNMPGRGGERLAVDPNKNISILYFGARSGHGLWKSTDYATGATWSNVTSTFTWGTGF 196  
QY 177 PTDITTYQSDIQGVVVAADFSSSLGQASKTIFVGVADPNNPVFWRDGGATWQAVPGA 236  
Db 197 SSST--YTSDDPVGIAWVTFDSTSGSGSATPRIFVGVADAGKSVFKSEADAGTAWVSGE 254  
QY 237 PT-GFIPKHGVPDPNVHLYIATNSGTPYDGSQGVVWKFVSQVTSQWTRISPVSTDTAN 295  
Db 255 PQYGFPHKGVLSPKELYIISYANGAGPYDGTNGTVHKYNTISGVWTDISP----TSLAS 311  
QY 296 DYFGYSGLTIDRQHPNTIMVATQISWPDITIIFRSTDCGATWTRIMWTSYVNSLRVYL 355  
Db 312 TYGYGGLSVLDIQVFGTLVVAALNCWMPDELIFRSTDSGATWSPIWENWGYPSINYYSY 371











A:Reference number: S74322; MUID:97061201; PMID:8905231

A:Accession: S75251

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-3972 <KAN>

A:Cross-references: UNIPROT:P73139; EMBL:D90904; GB:AB001339; NID:gl652225; PIDN:BAAL1716

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Superfamily: Synchocystis hypothetical protein slr1028

Query Match 4.7%; Score 189; DB 2; Length 3972;

Best Local Similarity 19.3%; Pred. No. 0.014;

Matches 166; Conservative 93; Mismatches 276; Indels 326; Gaps 42;

```
QY 13 IGGGFGVGVNNEGAGLIVRTDIGMYRDAANG-----RWIPLLDW- 57
DB 2458 VNGDGFADVISGGSPAGGVIFGNSTKOLL--DAALGTDDLIIISVENAQKVFALGDFD 2515
QY 58 -----VGNWNGYNG---VWSIAADPINTNKVAAVGMVMTNSWD-- 93
DB 2516 GDGLADFGVIDDQGNFFVLVSGPELGSQSLVIDSTLPNLSNFNQAWGVDFNGGYDDF 2575
QY 94 ----PNDGAILRSSDQGATWQITPILPKLGGMP-----GRGMGERLAVDPN-- 136
DB 2576 VLQGNSTIAVYGNANGTLTDSPLTF--GNFPLPSSFTGIDLNGNGIKIIVAGQPNLN 2633
QY 137 -----NDNLIYFGAPSGKGLWRST-----DSG-ATWSQMTNFPDVGTYIANP 177
DB 2634 PVNIGGFGGLQVFTYEAGNAVLOFTVNPNNASVTEASGLSSWQI-SFPNOYAAQAVP 2692
QY 178 TDTT-GY-----QSDIQGVW-----VAFDKSSS----- 200
DB 2693 SPATLDGMDYQAFYGINERISKDSYIYIQRSDGVSWENLVQVPLDSNGTPIDLKNLP 2752
QY 201 -SLGQASKTIFVGVADPNPNVFWSRDG-----GATWQAVP-----GAPT--GFIPHK 244
DB 2753 PSITANGTYLIGFTADNGQV-WVAEGVNTANSGLINAVPINQASNNGTPLVAFNDEL 2811
QY 245 GVF----DPNVHVIATNTGPGVDGSSGVWKFVSTGWTTRISPVSTDTANDYFGY 300
DB 2812 YVFFVKDASNDILYSSSNPG-----SSSG--WDGTSTVLTFSDVNAQTN-----PFL 2858
QY 301 SGLTIDRQHPNTIMVATQISWHPDTIIFRSTDGGATWTRI-----WD----- 342
DB 2859 SATWPEGLDGTFLAVA-----FRSNNSPATWVGLNNSDVNTNMGSAELTQVDA 2907
QY 343 -----WTSYPNRSRLRYL-----DISAEPWLTFGVQPNPPVPSPKL 378
DB 2908 NSQVSLTVVDGYTYLFFSTSTASASYATSDGLNWGDITLIPW-----DDGNL 2956
QY 379 GWMDE-----AMADFPNSDRMLYGTGATLYATNDLTKWDSGGQIH----- 419
DB 2957 GGVASILFNQSFILSNQNNESLLFAPSNSLFEPPNQASRW--GEQVRDIDGPNGDGIAD 3014
QY 420 ----IAPMVKGLBEETAVNDLISPPSGAPLISALGDL-----GGFTHADVAVPSTIPTSVP 472
DB 3015 LAVLAPGYRNLLQFPILDY-----PAINNLLGVFIYIEESGISVNDPDPVLAAPDL 3067
QY 473 TTGTSVDVYAEALPSILVRAGSDPP--SSQP-----NDRHVAFPSTDGGKNW-- 515
DB 3068 PQSTIFELLEITPTGVDNGDGFDDLLISAPLTPVIAQFPDVGNGDGVSWVFGGTHGT 3127
QY 516 -FGSEP-----GGVTTG-----GTVAASADG-SRP-VW 541
DB 3128 EYTANSPFGLGNLANNQTNNSQNFYGFVTTGLPRSQAGISISGGADVNGDVSDFALG 3187
QY 542 APGD-----PGQPVVAVGFGNSWAASQ-----VPAQAQIRSDR 576
DB 3188 APGNFNLNLYVLFGSDFTNQVNLQITGDDVMLGSPTEIFVAGQGDQDIYNTNGVDTVY 3247
QY 577 VNPFTFYALNSGTFYESTDGG 597
DB 3248 AGPNDFVTVTDTNFRRLDGG 3268
```

#### RESULT 8

S76412

hypothetical protein slr0408 - Synchocystis sp. (strain PCC 6803)

C:Species: Synchocystis sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 09-Jul-2004

C:Accession: S76412

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;

O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synchocystis

S.

A:Reference number: S74322; MUID:97061201; PMID:8905231

A:Accession: S76412

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-4199 <KAN>

A:Cross-references: UNIPROT:P74440; EMBL:D90915; GB:AB001339; NID:gl653604; PIDN:BAAL1815

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Superfamily: Synchocystis hypothetical protein slr0408

Query Match 4.6%; Score 186.5; DB 2; Length 4199;

Best Local Similarity 19.2%; Pred. No. 0.021;

Matches 198; Conservative 124; Mismatches 319; Indels 389; Gaps 54;

```
QY 9 SNAVIGGGFGVGVIV-----FNEGAGLIVRTDIGMYRDAANGRWIPLL 55
DB 2261 ADVASDNGFVIDGNLIGNPPTFTTTSQYIDTTPAILNGSNLYLAYKFGGNQ---IY 2317
QY 56 DWVGWNN-WGYNGVWSIAADPINTNKV-AAVGMVMTNSW-----DPNDGA-ILRSSDQGA 107
DB 2318 FTVSTNGQSWNSEVL---PQSAQTIFFPAIAFFNNVLYAVDGNNGNLNIITSQDQGQ 2374
QY 108 TWQITLPPKLGGMPG-----RG-MGERLAVDPNNNDNILFVGPAGSGKGLWRSTDSGA 159
DB 2375 TWN---APLALGTSSTPPTLFVYQGTLSLLFAANNSTSTVLOFYLNSSNEWIYANEIGS 2431
QY 160 TWSQMT-----NFPDVGTYIANPDTTTCGYQSQDIQGVWVAF----- 195
DB 2432 NOTAISAISATVLGDTLYLYKGGTRNTPSTLDYITSTN-----ADLSANDMSSSIPGP 2486
QY 196 ---DKSSSSILGQASKTIFVGVADPNPNV-FWSRDCGATWQAVPGAPTGF1---PHKGVFDP 249
DB 2487 VSSQGGPSLTNDGTNLYLYSVDSSNQNLNFVSSNGNLNWS--PQVITNNISQSPALAF-- 2543
QY 250 VNHVLIATNTGPGVDGSSGVWKFVSTGWTTRISPVST-----DTAN 295
DB 2544 ANNELYLS-----YPGQGS-QELNVTG-----PPLPFTGSLGNGSLVRFLGDVNG 2589
QY 296 DYFG--YSGLTIDRQHPNTIM-----VATQISWHPDTIIFRSTDGGATWTRIWDWTSY 346
DB 2590 DGFADVFSGGT---NAGAIIFGNSTKDLTLTASGSDELVI-----SV 2628
QY 347 PNRSLAYVL---DISAEPWLTFGV-----QBNPPVPSPKL 378
DB 2629 PNATLRDVISVGFNGDGIKDLGNGNPNFYVVLGNTSLGDLTKLTSITSSSPVPIVNOV 2688
QY 379 GWMDEAMADIPFNSDR---MLYGT-----GATLYATNDLTKWD----- 413
DB 2689 GGVTKSWAIGDYGNGDGYDDVLLWGDNGCNQVANGNSTGVLSNFTWIDYPTQTATTATVDLN 2748
QY 414 -----SGGQIHIAPMVKGLBEETAVNDLISPPSGAPLISALGDLGGFTHAD 458
DB 2749 SDGIPEIAGSDERKIAGQISTSGSFLLPTPTTSSVINTLAAANQLENIGDFFNGDGIAD 2808
QY 459 VTAVPSTIPTS-----PVFTTGTSDVYAEIN-----PSIIVRAGSDFPS 497
DB 2809 LAVIASNYAAATGEENPNLPNLYSRPQNGGVFIYFGNSNGLSNATAQPDVILAAFPPTNPS 2868
QY 498 SQ-----PNDRHVAFSTDGGKNW----- 515
DB 2869 GQISTVQLSRIAQAGDVNGDGFDDLLISSPYTVDAENNGGVFVVFSGDDNNQNPFDLQG 2928
```



QY 516 -----FOGSRGGVTTGGTVAASA-----DGSRF-----VWAPG----- 544  
Db 2929 LRANQSGSNPRFAIDGSPNSQAGIALNGGDDINGDFADFIIGAPGNNLQYNQIIVP 2988  
QY 545 -----DPGQPVVAVG-----FGNSWAASQGVPAQAIRSDRVN-----PKTFVAL 585  
Db 2989 IENGELSDDDKYSYILYLDGNQTIQMGGDQWQANQ-VMTN-QVATNWNNSRRPPEAVITGQ 3046  
QY 586 SNGTFYFSTGGVTFQ-----PVAAGLPSSGAVGVMPHVPKG 623  
Db 3047 SNGDIWTPYPGNQWQSGKLPABINELAVNMTSGNPKIITAGLGKG--GIEYY----- 3099  
QY 624 EGDLLAASGLYHSTGGSSW-SAITGVSSAVNVGFKSAPGSYPVAVFVVGTVIGGVTG 682  
Db 3100 NGSTWV--NNGPYQ-----GDGWSRAITQMA-----VQWGED--GS--PSQIVVGLDAGVI 3145  
QY 683 AYRSDDCGTTVLINDQHQYGNWQAITG-----DHANLRVRYGTN 725  
Db 3146 YVNTQ-----SGWRTIN-----NFGKSVTQLSVQWQBASNPNIVVGLDENSEV-QVYQGSN 3194  
QY 726 GRGIVYGDIG 735  
Db 3195 GVMTQFHDG 3204

RESULT 9  
T13256  
tail-host specificity protein homolog - Lactococcus lactis phage BK5-T  
C;Species: Lactococcus lactis phage BK5-T  
C;Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 09-Jul-2004  
C;Accession: T13256  
R;Boyce, J.D.; Davidson, B.E.; Hillier, A.J.  
Appl. Environ. Microbiol. 61, 4089-4098, 1995  
A;Title: Sequence analysis of the temperate Lactococcus lactis bacteriophage BK5-T and c  
A;Reference number: Z17646; MUID:96064422; PMID:8526523  
A;Accession: T13256  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1904 <BOY>  
A;Cross-references: UNIPROT:Q38319; EMBL:L44593; NID:g928826; PID:g928828; PIDN:AAA98579

Query Match 4.5%; Score 182; DB 2; Length 1904;  
Best Local Similarity 20.7%; Pred. No. 0.013;  
Matches 189; Conservative 83; Mismatches 260; Indels 380; Gaps 55;

QY 1 ATTPQYTWNSVAICGGGFVDGIVFNEGAPG-----ILYVRTDIDGMYRWDANGR 50  
Db 838 AKPSDYTWS--LIRGNDGKDGATGKGVAGKDGVGIGKTTVITYALSSSG-----TDKPTNG 891  
QY 51 W---IPLLDWGVNWNWYGVVSTAADPINTNKVAAVGMVNTNSWDPNDGAILRSDQGA 107  
Db 892 WTSQVPTL-----VKGQYLTWTKVT-----YTDS-----SSETG- 921  
QY 108 TWQITPLPFLKGNMPCRMGERLAVDPNNDNILYFGAPSGKGLWRSTDGATW--SOMTN 166  
Db 922 -YSYVI--AKDGNNGDGIAGKGVGKIKTIYAVGTSG-----TTAPASGNSQVFN 973  
QY 167 FP-----DVGTYIANPTDTTGYSDIQGVVWVAFDKSSSSSLGQASKTIFVGVADPNN 218  
Db 974 VPAGQFLWTKTVIYTDN--TSETGYSVAMGVKDGKDPGNGTN-----GIA----- 1020  
QY 219 PVFWSRDG-----GATWQAVPG---APTQ---FIP---HKGVP-----DP 249  
Db 1021 ----GKDGKIGIKATAITYQASPNGTAPTGTWSASVPPVAKGSFLWTRTIWYTDNTTET 1076  
QY 250 VNHVLYIATSNWGGPYGSSGDVWKFVSVTGWTFRISPVPTDITANDYFGYGLTIDRQH 309  
Db 1077 GYAVAYMGNTGNG--HDGFPG-----KQGTGKTTIT-----YAGSTSGTTP 1118  
QY 310 PNTIWAQISWPDITIFRSTDGATWTR--IWDWTSYPNRSRYVLDISAEPMLTFCVQ 368  
Db 1119 EN-----NGWTSYTP--TYAEGNLYWTKVIWYTD--NTS----- 1149

QY 369 PNPPVPSKLGWDEAMAI-----DPFNSDRMLYGTGATLYATNOLDTKWDSGQIHAPM 423  
Db 1150 -----ETGYSVAMGVKDGKDPGNN-----GTNGIAGKDG----- 1180  
QY 424 VKGLEETAVNDLISP-PSGAPLISALGDLGGFTHADVTAVP--STIFTSFVFT----- 473  
Db 1181 -KGIKATAITYQASPNGTAPT-----GTWSASVPPVAKGSFLWTRTIWYTDNTTTE 1231  
QY 474 TGTSVDVYAEIN-----PSIIVRAGSPDPSPQPNDRHVAFASTDGKKNWF 516  
Db 1232 TGYAVAYMGNTGNGHGDGPKDGTGKIKTITITYAGSTSGTTPNN-----GW- 1279  
QY 517 QGSEBPGVTTGGTVAASADG---SRFYW-----APGDGQPVVYAVG 555  
Db 1280 -----TSVTPTVAEGNLYWTKVTWYTDNTSETGYSVAMGVKDGKDP----- 1323  
QY 556 FGNSWAASQGVPAQAIRSDRVNPKTFYALNSGTFTYRSDGVTGTPQVPAAG----- 606  
Db 1324 -GNN--GTNGI--AGKDGKIGIKATAITYQASPNGTAPTGTWSASVPPVAKGSFLWTRTIW 1379  
QY 607 -----LPSSGAVGVN-----PHAVPKEGD-----LWLAASSGLYHSTNGSS-- 644  
Db 1380 TYTDNTTETGYAVAYMGNTGNGHGDGPKDGTGKIKTITITYAGSTSGTTPNNGWISTV 1439  
QY 645 -----W-----SAITGVSSA--VNVGFKSAPGSYPVAVFVVGTVIGGVTG-- 682  
Db 1440 PTVAEGNLYWTKVTWYTDNTSETGYSVAMGVKDGKDPGNN-----GT--NGIAGKD 1491  
QY 683 -----AYRSDDCGTTVLINDQHQYGNWQAIAI-----TGDHANLRV----- 720  
Db 1492 GKGIKATAITYQASPNGT-----APTGTWSASVPPVAKGSFLWTRTIWYTDNTTTE 1543  
QY 721 -----YIGTNG 726  
Db 1544 TGYAVAYMGNTG 1555

RESULT 10  
F70846  
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)  
C;Species: Mycobacterium tuberculosis  
C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C;Accession: F70846  
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;  
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A;Reference number: A70500; MUID:98295987; PMID:9634230  
A;Accession: F70846  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-2523 <COL>  
A;Cross-references: GB:AL021841; GB:AL123456; NID:g3261517; PIDN:CAAL17115.1; PID:e125111  
A;Experimental source: strain H37RV  
C;Genetics: PPE  
A;Gene: PPE

Query Match 4.3%; Score 175.5; DB 2; Length 2523;  
Best Local Similarity 21.8%; Pred. No. 0.047;  
Matches 175; Conservative 92; Mismatches 327; Indels 209; Gaps 38;

QY 10 NVAIGGGFVDGIVFNEGA--PGIYVRTDIDGMYRWDAN--GRWITPLLDWGVNNGYN 66  
Db 1522 NTGICNSGIASGLFNAGFNTGVVAGSYNTGSGFNAGQANTGGFNPGSVNTGLNTG-- 1579  
QY 67 GVWSTAADPINTNKVAAVGMVNTNSWDPNDGAILRSS--DQATWQITPLPFLKGNMGP- 124  
Db 1580 -----DINTG-----VANSGDVNTGAFISGNSNGAFWR-----GDYQGL 1614  
QY 125 RCMGERLAVDPNNDNILYFGAPSGKGLWRSTDGATWSQMTNFPDVGTVIANPTDTTGYQ 184



Db 1615 LGFSVRPAVLQ-----TPFLDLTLTGLGSVWIPADIPAIR 1652  
QY 185 SDIQGVVWV-APDKSSSSLGQ---ASKTIFVGVAD-----PNNPVFVRDGGATWQA 232  
Db 1653 PEFSANVAIDSTVPSIPQIDLAATVSVGLGPIVPHLIDIPRPVPTLNYLFGSQ--- 1709  
QY 233 VPGAPTGPVPHKGVDPVNHVLYATSNTPGPDGSSGVKMFVTSQVTRISVPSTD 292  
Db 1710 -PGGPKKIGPITGLPFTPIGLTPLALSIQIVIGASSQGTITAFANLFPSTVVTIDIP 1768  
QY 293 TANDYFGS-----GLTIDROHPNTI-----MVATQISWMPDTIIFRSTDDG 334  
Db 1769 LLASITGSEPDVDFPGGLTIPAMNPLSINLSCGTGAVTIPALITGIBPFDLVAHSTLG- 1827  
QY 335 ATWTRIMDWTSPNRSRLRYLDISABPWLTFGVQVNPVPVPGKLGWDE----- 383  
Db 1828 -----PVHILIDLPAVP--GFG-----NTTGAPSSGFFNSGAGGVSGFGNV 1866  
QY 384 -AMADPFN-SDRMVLYGTGATLYATNDLTKWDSGQIHIAPMVKLEETA VNDLISPPSG 441  
Db 1867 GAMVSGGNQAPSALLGGSGGVFNAGTL-----HSGVLYNFGSGMGLFNTSVLGL-----G 1917  
QY 442 AP-LISALGDLGFTHADVTAVPSTIFTSPVFTTGTSDVYAEINPSIIVRAGSFPSSOP 500  
Db 1918 APALVSGLGSGVQ-----QLSGLLASGTA-----LHQVLNFGGLADVGL-- 1957  
QY 501 NDRHAFSTDDGKNVFGQSEPGVTTGGTVAAADSRFV-WAP-----GD-----PGQPVVY 552  
Db 1958 --GNVGLGVDFNLGAGNVGFFVGGNIGNNVGLGVNGFGLGNSGLTGPLMGLG 2015  
QY 553 AVFGNSWAASQGVANQIRSDRVNPKTFYALSNGTF-----YRSTDGV 598  
Db 2016 NIGFNGAGSYNFGI-ANM-----GVNIGCFANTGSGNFGIGLTGDNLTGFGGFNTGSGNV 2069  
QY 599 TQPVAAGLPSSGAVGV-MFHAVPGKEGDLWLAASSGLVHSTNGSSSAITGVSSAVNV 657  
Db 2070 -----GLFNSGTVNGVFNFGTGN-----WGVFNSSGY-NTGIGNSGIASTGLFNAGGF 2117  
QY 658 GFGKAPSSSYPAVFGVGTGAYRSDDCGTTWVLINDQHOYGNWQA-----ITG 712  
Db 2118 NTGVNAGSYNTGTFNAGAN--TGFPNPGSVNTGWLNTGDTINTGVNSGDVNTGAFISG 2175  
QY 713 DHANLRRVYVGTNGRIVYGDIG 735  
Db 2176 NYSN-----GAFWRGDYQGLLG 2192

RESULT 11  
AE1905  
outer membrane secretion protein alr0791 [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp. PCC 7120  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Jul-2004  
C:Accession: AB1905  
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
DNA Rep. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AE1905  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-993 <KUR>  
A:Cross-references: UNIPROT:Q8YQ05; GB:BA000019; PIDN:BA072748.1; PID:g17130136; GSPDB:G  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: alr0791

Query Match 4.3%; Score 172; DB 2; Length 993;  
Best Local Similarity 19.6%; Pred. No. 0.022;  
Matches 168; Conservative 96; Mismatches 325; Indels 270; Gaps 44;

QY 45 DAAGRW-IPLLDWVGNWNGVNGVVSIAADPINTN-KV-----WAAVGMTNSW 92  
Db 58 DNSNGTWEVTLNNGTSWTTFGTPSLTAARLLPSNANTKIRFHPNANFSGTADINFY--AW 115  
QY 93 DPNDG-----ALRSSDQATWQITPL-----PFLGSGN--MPG----- 124  
Db 116 DQTTGTSSTANILAGKGTTFASTNYEGASITVTPVNDTAPTLLNAGSTPLPINEDAPL 175  
QY 125 -RGMGERLA-----VDPNNNDILYFGAPSGKGLMR--STDSGATWSQMTNFPDVGTY 173  
Db 176 ISNKGSLIADLVRLGISDSDDPYQIAVTGADNNGSQWYSLDGGVNLNFGNASDNSAT 235  
QY 174 IANPTDTTGYQSDIQGVWVAFAFKSSSSSQASKTIFVGVADPNPNVFWVRDGGATWQAV 233  
Db 236 LLLPSIR-----LYDGSLSLPTSQWMLKFGASPEVFLF-----PVGQTSLI 279  
QY 234 PGA-----PTGFIPHKGVDPVNVHLYIATSNTPGPDGSSGDVWVKFVTSQVTRISPVPS 290  
Db 280 TGGIQLNSSSIGSSG--YSNYSYAPILFNQAPPELDPVKGTTFISFDVKINGETHSD-DN 337  
QY 291 TDTANDYFCYGLTIDROHPNTIMVATOISWMPDTI-----IFRSTDGG 334  
Db 338 GGIQDRAGFSVIVTSDTKAI-----ELGFWDEIWAQTASPLFTHSTITERAFRNTT 393  
QY 335 ATWTRIMDWTSPNRSRLRYLDISABPWLTFGVQVNPV-----PSPK 377  
Db 394 VT-----RVHLVVENNTYKLFAPDSTPILSGNLRDYSAFNHSTAAPSP- 437  
QY 378 LGWMDENMAIDPFNSDRMLY-GTGAT-LYATNDLTK----- 411  
Db 438 ---ITSLPFPDPVETPNFLGDNLTTSQAASSNLQVELQTNTRVRFVFPNADYNGQANLT 493  
QY 412 ---WD-----SGQIHIAPMVKLEETA VNDLISPPSGAPLISALDGLGFTHADVTAVP 463  
Db 494 FRAMWDSNGVAGG-----TTGVNAAVN-----GNATAFSSNTLTA-- 528  
QY 464 STIFTSV-----FTTGTSDVYAEINPSII-----VRAGSFPSSQPNDRHVAFS 508  
Db 529 -SITVSPINNPITQTTGLDKLYGTANEDIIINGEGNDYLFGRAGNDLTDGEGNDYLF 587  
QY 509 T-----DGGK--NWFGQSEPGVTTGTTVAASADGSRFVWAPGDPGVVYAVGVFNSW- 560  
Db 588 TGNLTDGEGSDLLYGNEDNDIINGGVGNDLNG-----GTGDD-----ILRGGTGNDIY 638  
QY 561 ---AASQGVANQIRSDRVNPKTFYA-----LSNGTFYRST--DGGVTFQFVAAGLPSSG 611  
Db 639 TVDTVGDVIEENPNEGTDKNVSYISWTLGANLENLTLGNTIIDG--TGNELDNHIIINN 696  
QY 612 AVGVMFHAVPGKEGDLWLAASSG--LVHSTNGSSSAITGVSSAVNVGPKSAPGSYP 669  
Db 697 AV-----NRLEGGDNDWLIGKQNDILIGGNDRLNGETG-EDTLEGGLGND----- 744  
QY 670 AVFVVGTTIGVGTGAYRSDDCGTTWVLINDQHOYGNW-----GQAITGDHAN 716  
Db 745 -VVEIDSVGDV--IEADAGIDTVISSVD-----WTLGVLENLTLVGNQATLIGND 795  
QY 717 LRRVYVGTNGRIVYGDIG 735  
Db 796 LDNRITGNNAADNVLFGEAG 814

RESULT 12  
AE2254  
hypothetical protein alr3588 [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp. PCC 7120  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Jul-2004  
C:Accession: AB2254  
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
DNA Rep. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium An  
A:Reference number: AB1807; MUID:21595285; PMID:11759840







QY 642 -GSSNSALITGVSAVNVGFGKSAAGSPVPAVFVGTIGGVT-----GAYR-----SDDCG 690  
 Db 4776 TQGSY-LYTNVSNASN--WTQSALGGSQS-----GTTSGVAAMDYDWDGAVDVLVSKQS 4827  
 QY 691 TTWVLINDDOHQYGNWQAITGDHANLRVYVGTNGRGIVYGD 733  
 Db 4828 SVFLSRNTNTVSYG-----TSLHLRI-----TDPNGINVYGN 4860

RESULT 14  
 F90696  
 hypochetrical protein EC0542 [imported] - Escherichia coli (strain O157:H7, substrain R1)  
 C:Species: Escherichia coli  
 C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004  
 C:Accession: F90696  
 R:Hayaishi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;  
 Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Res. 8, 11-22, 2001  
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno  
 A:Reference number: A99629; MUID:2115231; PMID:11258796  
 A:Accession: F90696  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Cross-references: UNIPROT:O8X2T1; GB:BA000007; PIDN:BA033965.1; PID:gl3360000; GSPDB:G  
 A:Experimental source: strain O157:H7, substrain R1MD 0509952  
 C:Genetics:  
 A:Gene: EC0542

Query Match 4.2%; Score 168.5; DB 2; Length 5291;  
 Best Local Similarity 20.1%; Pred. No. 0.32;  
 Matches 190; Conservative 82; Mismatches 270; Indels 401; Gaps 46;

QY 27 GAGPILYVRTDIGMYRMDAANGRWIPLLDWGMNNGVYVIAADPINTNKVAAVG 86  
 Db 4186 GANFPAQISTDNGA-----TWVN-----VTVAADSLN-----4212

QY 87 MYTNSWPNDCGAILRSSDQATWQITPLPKLGGNPMRGMRGRLAVDPNN-----137  
 Db 4213 -----WSYVDGRTLTNGT--TTQVRRV--DLAGNV-GATSSQSAIDVTNPAQVLTIAS 4262

QY 138 -----DNILYF-----GAPSKGLWRSTDGATWSQNTNPDVGTIYIAN 176  
 Db 4263 ISTDGSSATPITSDMTLTLTGSLAGLASEVAQISLDGATWTLTTLTNGTQMTYTD 4322

QY 177 PDDTTG---YQSDIQGVVWVAFDKSSSLGQAKTIFVGVADPN---NPVFWSRDGGATW 230  
 Db 4323 RFLTDGSVYQVRV-----LDLAGNTGPPVSKTVVDVTINPTATPTIVSVTDDVGQRQ 4375

QY 231 QAVPGAPTGFIPHKGVFDPVNHVLYIATSNTPGPPYDG-----SSGDV-----272  
 Db 4376 GTLSSSQ-----ATDDTTPLLNGVLSPALASGEVYLYRNGLLGA 4416

QY 273 -----WKEF---VTSQGTW---RISFPVSTDTANDYFGVSGLTIDRPHNTIWTAT- 317  
 Db 4417 VTMWALNWTYSDSGLVSGATYSARVVDLAGNITSSSDP---VLTVDTSIPTTLAQITS 4473

QY 318 -----QISWN---PDT---325  
 Db 4474 QTRDTTPIISGVITAAALASQYVEVINGKTYTSEPGAVVDPAHNTWYVQLPDTAL 4533

QY 326 -----IIFRSTDDGATWTRITWDTSYNRSRLRYVL 355  
 Db 4534 TVSATAYTVTAQVKSAGNANNISNGVTVNAALIDYPTTWTASKTTAW---GLTYGL 4590

QY 356 DISAEFWLTFGVQPNPPVSPKLGWMDAMADPFSNDRMLYGTGATLYATNDLTKWDSG 415  
 Db 4591 D-SHGMMTVLANQQVWQSQSDTP-LTWSKTALT-----LYQSGNN-YATSSIAADYDRN 4638

QY 416 QQHIAPIWVKGLEETAVNDLISPPSGAPLIS---ALGDLGGFT-HADVTAVPSITFTSP 470  
 Db 4639 G-----TGDLFITRDDYGTGYINGFTNNGDGT-----PSSA 4669

QY 471 V-FTTGTSDVYAEALNPSIIVRAGSPDPSSQPNDRHVAFST--DGGKNWFGQSEPG-----522  
 Db 4670 IQVTGTLTWGSI-----VAFDKEGDGYLDFWIGDAGGPDSENT 4708

QY 523 -----GVTTCGTVAASADGSRFFWAPGDPGQPVVYAVGFGNSWAASQGVPAQAIRSDR 576  
 Db 4709 FLNNAGTFLVGNSTTSNGSGSATVGG-----AVTGYLSLNEGSGVDLNNDRIDL 4758

QY 577 V-----NPKTFVALS-----NGTFYRSTDDGGVTFFQPAAGLPSGAGVGMFHAVPQK-EG 625  
 Db 4759 VQHTYLNANYTLLSLINQNGTFFWQNTTNTFLSGAGSGAMSSSVMTWADFDDGDDM 4818

QY 626 DLWLAASSG-----LYHSTNG-----641  
 Db 4819 DLFLPASQGRANYGSLLENTNGVLGCPVAVGATATTYASQFSLAVDWNHGLMDIARIAQ 4878

QY 642 -GSSWSALITGVSAVNVGFGKSAAGSPVPAVFVGTIGGVT-----GAYR-----SDDCG 690  
 Db 4879 TQGSY-LYTNVSNASN--WTQSALGGSQS-----GTTSGVAAMDYDWDGAVDVLVSKQS 4930

QY 691 TTWVLINDDOHQYGNWQAITGDHANLRVYVGTNGRGIVYGD 733  
 Db 4931 SVFLSRNTNTVSYG-----TSLHLRI-----TDPNGINVYGN 4963

RESULT 15  
 A28452  
 proteoglycan core protein precursor, cartilage - rat  
 N:Alternate names: aggrecan  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 09-Jul-2004  
 C:Accession: A92623; A28453; A28454; A28455; A28456; A28457  
 R:Doerge, K.; Sasaki, M.; Horigan, E.; Hassell, J.R.; Yamada, Y.  
 J. Biol. Chem. 262, 17757-17767, 1987  
 A:Title: Complete primary structure of the rat cartilage proteoglycan core protein deduced  
 A:Reference number: A92623; MUID:88087070; PMID:3693370  
 A:Accession: A92623  
 A:Molecule type: mRNA  
 A:Residues: 1-2124 <DO>  
 A:Cross-references: UNIPROT:P07897  
 R:Doerge, K.; Sasaki, M.; Horigan, E.; Hassell, J.R.; Yamada, Y.  
 J. Biol. Chem. 263, 10040A, 1988  
 A:Reference number: A30069  
 A:Contents: annotation; revision to residue 698  
 R:Doerge, K.; Fernandez, P.; Hassell, J.R.; Sasaki, M.; Yamada, Y.  
 J. Biol. Chem. 261, 8108-8111, 1986  
 A:Title: Partial cDNA sequence encoding a globular domain at the C terminus of the rat  
 A:Reference number: A23835; MUID:86250698; PMID:2424893  
 A:Accession: A23835  
 A:Molecule type: mRNA  
 A:Residues: 1856-2124 <DO2>  
 A:Cross-references: GB:M13518; NID:G206104; PIDN:AAA41836.1; PID:G206105  
 R:Neame, P.J.; Christner, J.E.; Baker, J.R.  
 J. Biol. Chem. 262, 17768-17778, 1987  
 A:Title: Cartilage proteoglycan aggregates. The link protein and proteoglycan amino-ter  
 A:Reference number: A28453; MUID:88087071; PMID:3693371  
 A:Accession: A28453  
 A:Molecule type: protein  
 A:Residues: 20-37, 'W', '39-60, 'B', '62-64, 'X', '66-69; 70-83; 84, 89-148, 'L', '150-238, 'S', '240, 'A'  
 C:Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology; B  
 C:Keywords: glycoprotein  
 F:1-19/Domain: signal sequence #status predicted <SIG>  
 F:20-2124/Product: proteoglycan core protein #status predicted <MAT>  
 F:44-135/Domain: immunoglobulin homology <IMM>  
 F:170-247/Domain: link protein repeat homology <LNK1>  
 F:268-349/Domain: link protein repeat homology <LNK2>  
 F:504-581/Domain: link protein repeat homology <LNK3>  
 F:602-683/Domain: link protein repeat homology <LNK4>  
 F:1914-2034/Domain: link protein repeat homology <LNK5>  
 F:2041-2097/Domain: complement factor H repeat homology <PHD>  
 F:126, 239, 333, 387, 611, 667, 1842/Binding site: carbohydrate (Asn) (covalent) #status pred

Query Match 4.2%; Score 168; DB 2; Length 2124;



























```

QY 359 ABPWLTFGVQNPVPSP---KLGWDEAMADFPNSDRMLYGTGATLYATNDLTWKDWSG 415
D 379 LAPW---IGNIVDVTPGNLQICWMMESLSIDFFDSNHMLYGTGETIYSGRDLWKDWSA 434
QY 416 GQIHIAPMVKGLETAVNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTG 475
D 435 HNVTIKSLADGVETSTQALISPPSGPLVSAIGDVGVGHQSLTRAPSAQFTNPKNWTS 494
QY 476 TSVDYAEINPSIIVRAGSDFSSQPNDRHVAFTDGGKNWFGQSGEPGGVTTGGTVAAASD 535
D 495 ADIDFAGNPNENVRIGTGDS---GKQVAISSDYGVTVNQHGAPDNVGGKVALSAD 551
QY 536 GSREVPWAPDGPQPVVYAVGFSGNSWAASQGVPAANAQIRSDRNPRTFYALSNGTFYRSTD 595
D 552 ADIILWRTNGNG-----VMVSRNQATENI FYGASGKTFYVSTD 589
QY 596 GGVTFQPVAAGLPSSGAVGVMFHVPKGBGLWLAASSGLYHSTNGSGSAITGVSSAV 655
D 590 NGKTFSAHGSLGATSVDITVH--PSVSGDIWASTDKGLFHSFSDSGATFSAISGITOAW 647
QY 656 NVGFGKSPAGSSYPAVVGTGIGVGTGAYRSDDCGTTWVLINDDOHQYG-NWQQAITGDH 714
D 648 GVALGAPRSTGGYPVFAAANYGVEAYFRSDRGVNVWVKINDAAHGFGAASANCMAADP 707
QY 715 ANLRVYVGTNGRIGVYGDIGGA 737
D 708 RVYGRVYVGTNGRIGVYGDVAGS 730

RESULT 8
O74170
ID O74170 PRELIMINARY; PRT; 856 AA.
AC O74170;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Avicelase III.
GN Name=avicii;
OS Aspergillus aculeatus.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5053;
RN [1]
RP SEQUENCE FROM N.A.
RA Arai M., Takada G., Kawaguchi T., Sumitani J.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RL EMBL; AB015511; BAA29031.1; -.
DR PIR; T00349; T00349.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR000254; CBD_fungal.
DR InterPro; IPR002860; Glyco_hydro_BNR.
DR InterPro; IPR011040; Sialidase.
DR Pfam; PF00734; CBM 1; 1.
DR ProDom; PD001821; CBD_fungal; 1.
DR SMART; SM00236; fCBD; 1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
SQ SEQUENCE 856 AA; 89820 MW; BE085983AF60ED76 CRC64;

```

Query Match 41.6%; Score 1680; DB 2; Length 856;

Best Local Similarity 46.2%; Pred. No. 1.2e-87;

Matches 346; Conservative 113; Mismatches 258; Indels 32; Gaps 17;

1 ATTQPYTWSNVAI--GGGFGVDGIVFNECAPGLYVVRTDIGNMYRWDAAANGWIPLLDWWG 59

21 AASQAYTKNVVTVGGGGFTGIVNPSAKGVAVARTDIGAYRLN--SDDTWTPLMDWVG 79

60 ---NNWNGYGVGSAADPINTNKVAAVGYMTNSWPDNGAILRSSDQGAQTWTPLPF 116

80 NDTWHD--GIDALATDPDVTDRVYVAVGYMTNEDWNVGSILRSTQGDWTETKLPF 136

```

QY 117 KLGNNMPCRGMERLAVDPNNNDNIIYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIAN 176
D 137 KVGGNMPCRGMERLAVDPNNKNSIIYFGARSCHGLWKSTDYDGTATWSNVTSFTWTGTYFQD 196
QY 177 PTDITGYQSDTQGVVWVAFDSSSLGQASKTIIFGVADPNPNPVFWRDGGATWQAVPGA 236
D 197 SSST--YTSDPVGIAWVTFDSTSGSGSATPRIFGVADAGKSVFKSESDAGATWAWVSGE 254
QY 237 PT-GFIPKHGVPDPVNHVLIATSNTPGYPDGGSSGDVMKFSVTSQGTWTRISPVPSPTDAN 295
D 255 POYGFELPHKGVLSPEKTIYISYANGAGPYDGTNGTVHKYNTISGVWTDISP---TSLAS 311
QY 296 DYFGYSLGTLIDRQHPNTIMVATQISWMPDTIIFRSTDGGATWTRINWTSYPNRSLRYVL 355
D 312 TTYVGGLSVLDLQVPGTLMVAALNCWPFDELIFRSTDSGATWSPIEWENGYPSINYYSY 371
QY 356 DISAPFWLTFGVQPNP--PVPSPKLGWDEAMADFPNSDRMLYGTGATLYATNDLTWKDWS 414
D 372 DISAPWIQDTTSTDQFPV--RVGMVVEALADFPDSNHMLYGTGLTVYGGHDLTNWDS 428
QY 415 GQIHIAPMVKGLETAVNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTT 474
D 429 KINVTVKSLAVGIEEMAVLGLITPPGFPALLSAVGDDGGFYHSDLDRAFPNQATHTPYGT 488
QY 475 GTSVDYAEINPSIIVRAGSDFSSQPNDRHVAFTDGGKNWFGQSGEPGGVTTGCTVAASA 534
D 489 TNGIDYAGNKPNSIVRSGASD--DYPT---LALSSNFGSTWYADYAASTSTGTGAVALSA 543
QY 535 DGSRFVWAPDGPQPVVYAVGFSGNSWAASQGVPAANAQIRSDRNPRTFYALSNGTFYRST 594
D 544 DGDITVLLMSSTSGALVSKSQG---TLTAVSSLPSPGAVIASDKSDNTVYFGSGSAGAIYVSK 600
QY 595 DGGVTFQPVAAGLPSSGAVGVMFHVPKGBGLWLAASSGLYHSTNGSGSSWSAI--TGVSS 653
D 601 NTATSTFTKTVS--LGSSTTVNAI--RAHPSIAGDVWMASTDKGLWHSTYDGTFTQIGSVTA 658
QY 654 AVNVGFGKSPAGSSYPAVVGTGIGVGTGAYRSDDCGTTWVLINDDOHQYGNWQA--IT 711
D 659 GWSFGFGKASSTGYSVVIYGPFTIDGAAGLPKSEDAGTWQVTSIDASHGFGS--GSANVNV 717
QY 712 GDHANLRVYVGTNGRIGVYGDIGAPSG 740
D 718 GDLQTYGRVFRGHERPCHLLRQSRPAG 746

RESULT 9
Q7Z9M8
ID Q7Z9M8 PRELIMINARY; PRT; 838 AA.
AC Q7Z9M8;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Cel74a.
GN Name=cel74a;
OS Trichoderma reesei (Hypocrea jecorina).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
OX NCBI_TaxID=51453;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=QM6a;
RA Foreman P.K., Brown D., Dankmeyer L., Dean R., Diener S.,
RA Dunn-Coleman N.S., Goedegebuur F., Houfek T.D., England G.J.,
RA Kelley A.S., Meerman H.J., Mitchell T., Mitchinson C., Olivares H.A.,
RA Teunissen P.J.M., Yao J., Ward M.;
RT "Transcriptional Regulation of Biomass-Degrading Enzymes in the
RT Filamentous Fungus Trichoderma reesei.";
RL J. Biol. Chem. 0.0-0(2003).
DR EMBL; AY281371; AAP57752.1; -.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR000254; CBD_fungal.

```







```

Qy 361 PWTFTGVPQNPVPSPKLGMWDEMAIDPFRNSRMLYCTGATLYATNDLTWKDGGQIHI 420
Db 374 PWITGFLSQ---DTKHLGMMIESLEINPLDSHLYATGLTVYGGHDLTKWDIVHNVTI 430
Qy 421 APWKGLBEETAVNDLISPPSGAPILSALDGGFT---HADVTAVPTIFTSPVFTGTS 477
Db 431 QSLAVGIEEMAVLGLASAPGSELLAAVGDGFTFKSSDGLGTSPTKPTWMTPOWASSAD 490
Qy 478 VDVAELNPSIIVRAGSPDPSPQPNDRHVAFTSDCGKWFQGSSEPGVTTGCTVAASADGS 537
Db 491 VDVAENKPNVIRIGSSGQAQ-----VAVSSDGGASWHAHNGTDTTKSSGTVAISADAD 545
Qy 538 RFVWAPGDPGPVYVYVGFNGSMAASQGPANAIQIRSDRVNPKTFYALS-NGTFYRSTDG 596
Db 546 TIWSSGTSQ--VVRSQNG-QTFAVASLPSGAVIASDKRNTVYAGSITGTFTYRSTDT 602
Qy 597 GVTQPPVAAGLPSSGAV-GWVFHAPGKEGDLWLAASGLYHSTNGSSSAITGVSSAV 655
Db 603 GATFTAVSDALGSAKAVRDIVAH--PTVAGELYVSTDAGIFRSTDFGVSTFKLSGLTNTQ 660
Qy 656 NVGFGKSNPSSYPVAVFVGTGIGTAYRSDCGTWTWVLNDDHQQYGNWQA---ITG 712
Db 661 SVSLGVGTSQSGW-NLYAFGTGANGNKLYASADAGATWADVQGAQ-GFGSIAVASSKVAG 718
Qy 713 DHANLRVYIGTNGRGIYVIGDI---GGAP 738
Db 719 SANVPGQYVYVGTNGRGIYVYQAQVSSTGAP 747

RESULT 11
Q82K30
ID Q82K30 PRELIMINARY; PRT; 739 AA.
AC Q82K30;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 25, Last annotation update)
DE Putative glycosyl hydrolase.
GN OrderedLocusNames=SAV2574;
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis.";
RL Nat. Biotechnol. 21:528-531 (2003).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948; DOI=10.1073/pnas.211433198;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osomoe T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220 (2001).
DR EMBL; AP005031; BAC70285.1; -.
DR InterPro; IPR002860; Glyco_hydro_BNR.
DR InterPro; IPR011040; Sialidase.
DR InterPro; IPR006311; Tat.
DR Pfam; PF02012; BNR; 8.
DR TIGRFAMs; TIGR01409; TAT_signal_seq; 1.
KW Complete proteome.
SQ SEQUENCE 739 AA; 78399 MW; 8E3B0F3D0F5B3D28 CRC64;

```

Query Match 34.8%; Score 1405; DB 2; Length 739;  
 Best Local Similarity 40.8%; Pred. No. 4.9e-72;

```

Matches 308; Conservative 104; Mismatches 257; Indels 86; Gaps 21;
Qy 6 YTSNVAIGGGFVDFGIVFNEGAGILYVRDIDIGMYRWDAAANGRWIPLLDVWGNWNGY 65
Db 42 YRWRNAVIGGFTVGTGLVLFHPSVRGLAYARTDGGAYRWDARGARTPLIDHLGWDWNL 101
Qy 66 NGVVSIAADPINTNKVAAVGMVYNSWDPNDGAILRSDQAGATQIOTPLPFLKGGNMPGR 125
Db 102 LGVEAMAVDPTHPDRLYLAVGTYAQSWAGN-GAVLRSERDGTATRTDITVLKGGNEDGR 160
Qy 126 GMGERLAVDPNDNLLYFGAPSGKGLWFSTDSGATWSQNTNEPDVGTIYANPTDITGYQS 185
Db 161 GAGERLLVDPDRSDTLWLT-RHDLGLKSTDRGATWAAATAPP-----AKANS 207
Qy 186 DIQGVVWVAFDKSSSILGOASKTIIFGVADPNP-----VFWSRGGATWQAVPGAPTGF 240
Db 208 SQGVVF-----LVNAGRTVYAGWDGDTSGTANLYRTAD-GTTWGAIVGPRSGT 257
Qy 241 ---IPHKGVDFPNVHLYIATSNCTGPGYDGSQVWKFVSVTGWTTRISPV-----PSTDT 293
Db 258 SAKVPLRAAYDTHTRRELYVYDAPGPGQSDGSVHKLRTATGTWTEVPVKPGGTTSDG 317
Qy 294 ANDFYGSLGIDRQHPNTIIVATQISWNPDTIIIFRSTDCGATWTRINDWTSYFNRLRY 353
Db 318 SADTFAYGVAVDARRPGTLLVSTNNRWADGTVFRSTDCGRTWTSKLD-----AA 368
Qy 354 VLDISAEPWLTFGVQPNPVPSPKGLWMDMAIDPFNSDRMLYGTGATLYATNDLTWKD 413
Db 369 VFDVSETFLDWGDD-----KPKFGWIIQALAVDPYDSQHVYVYGTGATLYGTRDLKRW- 421
Qy 414 SGQQTIIAPVMVKLEETAVNDLISPPSG-APLISALGDLGGFTHADVTAVPST-IFTSPV 471
Db 422 -----APRIRGLESAVRQLISPPVGEAHLISGLDGIQVVMYHERLTASPSRGMATNPV 474
Qy 472 FTTGTSVDYAEINPSIIVRAGSDFSSQPNDRHVAFTSDGKNW--FQSEPGGVTTGTT 529
Db 475 FGSATGLAQAAARPAYVVRVTGMDHGNG-----AYSHDGGRTWAPFEAQPIDAKDAPG 528
Qy 530 VAASADGSRFYWA-----PGDPQPVVYA-----VGFNSWAAASQGVPANAIQIRSDRVNPKT 581
Db 529 IATSADGGTLLWSFVHWDG-----TTYAHRSTDNGASWSEVSPFKGATPVADPADPTR 583
Qy 582 FYA--LSNGTFFYRSTGGVTFQPVAAAGLPSSGAVGVMFHVPKKGDLWLAAS-SGLYHS 638
Db 584 FYAYDFDNGLYASTDSGRSFTARAGLP-SCDSQFKLVAAPGRSGLWLSAKWNGLYRS 642
Qy 639 TNGSSWSAITGVSSANVNVGFKSAPGSSYPVAVFVGTGIGTAVRSDDCGTTWVLND 698
Db 643 TDGGDTFARIDSCWASYTLGFGKAADGADYPAIYQVGSSTETITAVYRSDDAARTWVRIND 702
Qy 699 DQHQYGNWQAITGDHANLRVYIGTNGRGIYVGD 733
Db 703 DAHQWGTGEEAVVGDPRIGHRVYLTATNGRGIQYGE 737

```

RESULT 12

```

Q8PLM5
ID Q8PLM5 PRELIMINARY; PRT; 751 AA.
AC Q8PLM5;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Cellulase.
GN Name=celA; OrderedLocusNames=XAC1770;
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=2022145; PubMed=12024217; DOI=10.1038/417459a;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,

```



RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,  
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,  
RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,  
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,  
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,  
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,  
RA Katayama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,  
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,  
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezsa R.I.D.,  
RA Trindade dos Santos M., Truffi D., Teai S.M., White F.F.,  
RA Setubal J.C., Kitajima J.P.;  
RT "Comparison of the genomes of two Xanthomonas pathogens with differing  
RT host specificities";  
RL Nature 417:459-463 (2002).  
DR EMBL; AB011809; AAM36634.1; -;  
DR InterPro; IPR002860; Glyco\_hydro\_BNR.  
DR InterPro; IPR011040; Sialidase.  
DR Pfam; PF02012; BNR; 7.  
KW Complete proteome.  
SQ SEQUENCE 751 AA; 81259 MW; D4134D96285E599D CRC64;  
  
Query Match 28.8%; Score 1162.5; DB 2; Length 751;  
Best Local Similarity 35.7%; Pred. No. 3.4e-58;  
Matches 268; Conservative 115; Mismatches 306; Indels 61; Gaps 20;  
  
QY 2 TTPYTSNVALTGGGFFVDGIVFNEGAPGILVYRTDIGMYRWDAAANGRWIPLLDVGWGN 61  
DB TTPYTSNVALTGGGFFVDGIVFNEGAPGILVYRTDIGMYRWDAAANGRWIPLLDVGWGN 61  
QY 43 TPDYQVHVSVALGGGFFVGVLFHFAERDLAVARTDVGAYNDAQOQVWALTDLGAD 102  
DB TPDYQVHVSVALGGGFFVGVLFHFAERDLAVARTDVGAYNDAQOQVWALTDLGAD 102  
QY 62 NMGYNGVWSIAADPINTNKKVAAGVMTNSWDPNGAILRSSDQCATWQITPLPKLGN 121  
DB NMGYNGVWSIAADPINTNKKVAAGVMTNSWDPNGAILRSSDQCATWQITPLPKLGN 121  
QY 103 DNLMLGIDAFADPADPNALYLAAGTYMERAGN-AAVLRSPDRGQKQFERADLPFKLGN 161  
DB DNLMLGIDAFADPADPNALYLAAGTYMERAGN-AAVLRSPDRGQKQFERADLPFKLGN 161  
QY 122 MPORGGERLAVDPNDNLLYFGAPSGKGLWSTDSGATWSQNTNPPDVGTTIANTPTDT 181  
DB MPORGGERLAVDPNDNLLYFGAPSGKGLWSTDSGATWSQNTNPPDVGTTIANTPTDT 181  
QY 162 QLGRANGERLAVDPHDGRVLLGSRDA-GLWRSDDRGAHARVEGPPADALAGATARNHV 220  
DB QLGRANGERLAVDPHDGRVLLGSRDA-GLWRSDDRGAHARVEGPPADALAGATARNHV 220  
QY 182 GYQSDIQGVVWAFQKSSSLGQASKTIFVGVADPNNVPFVSRDGGATWQAVPGAPTGI 241  
DB GYQSDIQGVVWAFQKSSSLGQASKTIFVGVADPNNVPFVSRDGGATWQAVPGAPTGI 241  
QY 221 GSEQAV-GIAFVFEAASGHAGAATPRIYGVSTAGTSLYVSEDAGRTWSAVAGPKGLR 279  
DB GSEQAV-GIAFVFEAASGHAGAATPRIYGVSTAGTSLYVSEDAGRTWSAVAGPKGLR 279  
QY 242 PHKGVPDPNVHL-----YIATSNTPGYPDGSSGVWKFVSTGTSWTRISPFVSTDTA 294  
DB PHKGVPDPNVHL-----YIATSNTPGYPDGSSGVWKFVSTGTSWTRISPFVSTDTA 294  
QY 280 P-----SHVGHNAGQWLYSYGDRPGPDLWAGGALWNYDATQGRWEISPIQPAT- 330  
DB P-----SHVGHNAGQWLYSYGDRPGPDLWAGGALWNYDATQGRWEISPIQPAT- 330  
QY 295 NDYFGYGLTIDRQHPNTIMVATQISWPDFTIIFRSTDDGATWTRIDWTSYPNRSLRV 354  
DB NDYFGYGLTIDRQHPNTIMVATQISWPDFTIIFRSTDDGATWTRIDWTSYPNRSLRV 354  
QY 331 GDGFGWGAVADPQHPQVLLASTFRRTPRDEVFRSGDGRSVVPLAAAF----- 382  
DB GDGFGWGAVADPQHPQVLLASTFRRTPRDEVFRSGDGRSVVPLAAAF----- 382  
QY 355 LDISAEPWLTFGVQNPFPVPSKLGWDEAMATDPNSDRMLYGTGATYATNDLTKWDS 414  
DB LDISAEPWLTFGVQNPFPVPSKLGWDEAMATDPNSDRMLYGTGATYATNDLTKWDS 414  
QY 383 -DHSAAPW-TAHATPH-----WIG-ALADPPDSNHTATPVGTGIGWASRLQTFSE 430  
DB -DHSAAPW-TAHATPH-----WIG-ALADPPDSNHTATPVGTGIGWASRLQTFSE 430  
QY 415 GGOIHAPVKGLEETAENVDLISPPSGAPLISALGDLGGFTHADVAVPSTFTSPVFTT 474  
DB GGOIHAPVKGLEETAENVDLISPPSGAPLISALGDLGGFTHADVAVPSTFTSPVFTT 474  
QY 431 QQPLQWFWQDRGLEETVPLDLSPMAGALLSALGIDIGFRH-DALDTAQLQVLGRLTN 489  
DB QQPLQWFWQDRGLEETVPLDLSPMAGALLSALGIDIGFRH-DALDTAQLQVLGRLTN 489  
QY 475 GTSVDYAEINPSIIVRAGSPDPSSQPNDRHVAFTSDGKRWFO-GSEPGVTTGGTVAAS 533  
DB GTSVDYAEINPSIIVRAGSPDPSSQPNDRHVAFTSDGKRWFO-GSEPGVTTGGTVAAS 533  
QY 490 GESIDGAGQAPQVWVSGTV-RDRRNEIRALYSQDGTWAAAFASEPRGQAGTIAIA 548  
DB GESIDGAGQAPQVWVSGTV-RDRRNEIRALYSQDGTWAAAFASEPRGQAGTIAIA 548  
QY 534 ADGSRFWAPGDPQGVVAVG-FGNSWAASQGVPAQAQIRSDRVNPKTFYALS--NGTF 590  
DB ADGSRFWAPGDPQGVVAVG-FGNSWAASQGVPAQAQIRSDRVNPKTFYALS--NGTF 590  
QY 549 ADASQVWVVDQGG-----VMRTGDFGRWQVQGLPDTAVVAVADRVAQRWYAADRVSGRL 605  
DB ADASQVWVVDQGG-----VMRTGDFGRWQVQGLPDTAVVAVADRVAQRWYAADRVSGRL 605  
QY 591 YRSTGSGVTFQPA--AGLPSSGAVGVMFHVPKGGDLWLAASGLYHSTNGSSWS-- 646  
DB YRSTGSGVTFQPA--AGLPSSGAVGVMFHVPKGGDLWLAASGLYHSTNGSSWS-- 646  
QY 606 YESSDGAASPRDTGQQVGSFARDE-----RARQLRDPDRAGVYVYASFTLGVMRWQNG 660  
DB YESSDGAASPRDTGQQVGSFARDE-----RARQLRDPDRAGVYVYASFTLGVMRWQNG 660  
QY 647 ---AITGVSSAVNVGFKSAPGSSYPFVAVFVVGTTGGVTAIRSDGCGTTLVNLINDQHQY 703  
DB ---AITGVSSAVNVGFKSAPGSSYPFVAVFVVGTTGGVTAIRSDGCGTTLVNLINDQHQY 703

DB 661 QLRTLSKDEARSLGIGKALRAGAPPALYLAGRVAGVDGIFRSDDGEGHWRINDDAHRP 720  
QY 704 GNMGAQTGVDHANLRRVVIGTNGRIGVYGD 733  
DB 721 GK-PYSVTGDPRIAGRVVFATGGRGIFYGD 749  
  
RESULT 13  
QWYEL  
ID Q9WYEL PRELIMINARY; PRT; 707 AA.  
AC Q9WYEL;  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
DE Endoglucanase, putative  
GN OrderedLocusNames=TW0305;  
OS Thermotoga maritima.  
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.  
OX NCBI\_TaxID=2336;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=WSB8 / DSM 3109 / ATCC 43589;  
RX MEDLINE=99287316; PubMed=10360571; DOI=10.1038/20601;  
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,  
Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
McDonald L.A., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,  
Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.L.,  
Heidelberg J.F., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,  
Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;  
RT "Evidence for lateral gene transfer between Archaea and Bacteria from  
RT genome sequence of Thermotoga maritima";  
RL Nature 399:323-329 (1999).  
DR EMBL; AB001712; AAD35393.1; -;  
DR PIR; F72393; F72393.  
DR TIGR; TW0305; -;  
DR InterPro; IPR002860; Glyco\_hydro\_BNR.  
DR InterPro; IPR011040; Sialidase.  
DR Pfam; PF02012; BNR; 9.  
KW Complete proteome.  
SQ SEQUENCE 707 AA; 79496 MW; 8CD8743CBDA6A99F CRC64;  
  
Query Match 28.8%; Score 1160.5; DB 2; Length 707;  
Best Local Similarity 35.2%; Pred. No. 4.1e-58;  
Matches 264; Conservative 123; Mismatches 275; Indels 89; Gaps 26;  
  
QY 6 YTSNVAIGGGFVDGIVFNEGAPGILYVYRTDIGMYRWDAAANGRWIPLLDVGWGNWGY 65  
DB FEWKSVEINGGVFVPGIIFHPASPLLYARTDVGLYRWDEETKRWKQLDFLRRDQSDY 80  
QY 66 NGVVSIAADPINTNKKVAAGVMTNSWDPNDGAILRSSDQCATWQITPLP---FKLGNM 122  
DB NGVVSIAADPINTNKKVAAGVMTNSWDPNDGAILRSSDQCATWQITPLP---FKLGNM 122  
QY 81 MGVLVALDPSDPKRIYAMTKYTDW-AGYGAILI SEDYGETWTIVLNDKYGIKVGNE 139  
DB MGVLVALDPSDPKRIYAMTKYTDW-AGYGAILI SEDYGETWTIVLNDKYGIKVGNE 139  
QY 123 PGRGGERLAVDPNDNLLYFGAPSGKGLWRSSTDGATWSQNTNPPDVGTTIANTPTDT 182  
DB PGRGGERLAVDPNDNLLYFGAPSGKGLWRSSTDGATWSQNTNPPDVGTTIANTPTDT 182  
QY 140 DGRNAGERLQVDNFSVLFMGT-TKYLKWSDFGKWKVDSFPST----- 186  
DB DGRNAGERLQVDNFSVLFMGT-TKYLKWSDFGKWKVDSFPST----- 186  
QY 183 YQSDIQGVVWAFQKSSSLGQASKTIFVGVADPNNVPFVSRDGGATWQAVPGAPTGI 242  
DB YQSDIQGVVWAFQKSSSLGQASKTIFVGVADPNNVPFVSRDGGATWQAVPGAPTGI 242  
QY 187 -----SVTFVLFDEKSGEKSPTPRIFVGCSEPKG-IFVTEDGTTWNLPLNDLIP 239  
DB -----SVTFVLFDEKSGEKSPTPRIFVGCSEPKG-IFVTEDGTTWNLPLNDLIP 239  
QY 243 HKGVFDPVNHVLYIATSNTPGYPDGSSGVWKFVSTGTSWTRISPVSTDTANDYFGYSG 302  
DB HKGVFDPVNHVLYIATSNTPGYPDGSSGVWKFVSTGTSWTRISPVSTDTANDYFGYSG 302  
QY 240 LRCKIH--DGILVYVLSNALGPNAGTRGAVMKVIADQKWDYVTPMKGD-----FGYCG 291  
DB LRCKIH--DGILVYVLSNALGPNAGTRGAVMKVIADQKWDYVTPMKGD-----FGYCG 291  
QY 303 LTIDRQHPNTIMVATQISWPDFTIIFRSTDDGATWTRIDWTSYPNRSLRVLDISAEPW 362  
DB LTIDRQHPNTIMVATQISWPDFTIIFRSTDDGATWTRIDWTSYPNRSLRVLDISAEPW 362  
QY 292 IDVQE---NVVISTLDRWYPHDEIFISLNGGETWRPLLEKANF-----DINKAPW 339  
DB IDVQE---NVVISTLDRWYPHDEIFISLNGGETWRPLLEKANF-----DINKAPW 339  
QY 363 LTFGVQNPVPSPKLGWDEAMATDPNSDRMLYGTGATYATNDLTKWDSGQGIHAP 422  
DB LTFGVQNPVPSPKLGWDEAMATDPNSDRMLYGTGATYATNDLTKWDSGQGIHAP 422  
QY 340 IK---DLNP-----HWISD-VKIDPFDNRAIFTTGYGVWVYELKKSPEG---MGK 384  
DB IK---DLNP-----HWISD-VKIDPFDNRAIFTTGYGVWVYELKKSPEG---MGK 384



[illegible]

RESULT 14

QIDP9U5	PRELIMINARY;	PRT;	751 AA.
ID			
AC	QBP9U5		
DT	01-OCT-2002 (TtEMBLrel. 22, Created)		
DT	01-OCT-2002 (TtEMBLrel. 22, Last sequence update)		
DT	01-MAR-2004 (TtEMBLrel. 26, Last annotation update)		
DE	Cellulase.		
GN	OrderedLocusNames=XCC1752;		
OC	Xanthomonas campestris (pv. campestris).		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;		
OC	Xanthomonadaceae; Xanthomonas.		
OX	NCBI_TaxID=340;		
RN	[1]		
RP	SEQUENCE FROM N. A.		
RC	STRAIN=ATCC 33913 / NCPBPB 528;		
RX	MEDLINE=22022145; PubMed=12084217; DOI=10.1038/417459a;		
RA	da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,		
RA	Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,		
RA	Alves L.M.C., do Amaral L.M.C., Bertolini M.C., Camargo L.E.A.,		
RA	Camarotte G., Cannavan F., Cardozo J., Chambergro F., Ciapina L.P.,		
RA	Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,		
RA	Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,		
RA	Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,		
RA	Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,		
RA	Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,		
RA	Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,		
RA	Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,		
RA	Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,		
RA	Spínola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,		
RA	Trindade dos Santos M., Truffi D., Tsai S.M., White P.F.,		
RA	Setubal J.C., Kitajima J.P.;		
RT	"Comparison of the genomes of two Xanthomonas pathogens with differing		
RT	host specificities.;		
RL	Nature 417:459-463(2002).		
DR	EMBL; AE012276; AAM41043.1; -		
DR	InterPro; IPR002860; Glyco_hydro_BNR.		
DR	InterPro; IPR000169; Pept_Cys_acSite.		
DR	InterPro; IPR011040; Sialidase.		
DR	Pfam; PF02012; BNR; 7.		
DR	PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.		
KW	Complete proteome.		
SQ	SEQUENCE 751 AA; 81140 MW; 737856AA0B5698A4 CRC64;		

Query Match 28.4%; Score 1145; DB 2; Length 751;

Best Local Similarity 35.0%; Pred.No. 3.4e-57;

Matches 265; Conservative 121; Mismatches 295; Indels 78; Gaps 21;

Qy	1	ATTOPYTSNSVAICGGGFVDGIVFNEGAPGLIYVTRDIGMYRWDAANGRWITPLLDWUGH	60
Db	41	ATSGFYQWRSVAICGGGFVTGLVFLHFAERGLAYARTDVGGAIRWDAQAQQMTALTDWLGA	100
Qy	61	NNWGYGVSTFAADPINTNKVMAAVGMTNSWDPNDGAILRSSDQATWQITPLPFKLG	120
Db	101	DDMWLMGIDAPVDPADADALYLAAITWHERAGN-AAVLSFNGRIFERADLDPFKLG	159
Qy	121	NMPGRGMRGLAVDPNNNDNIIYFGAPSGKGLWRSTDGATQMSQMTNFFPDVGYIANTPTD	180
Db	160	NQLGRANGERLAVDPHGRVLLGSRDA-GLWRSDDRGAHWAKVASFDDAALAGATARNH	218
Qy	181	TCYQSDTQGVVWVAFDKSSSLGQASKTIFVGVADPNPNPFWRSRGGCATWQVAGNPTGF	240
Db	219	VGREQAV-GIAFVVFDAAAGNTGTPTRIIYGVGVSTQSLTVSESDAGRSWAPVAGQPRGL	277
Qy	241	IPHKGVDPVNVNHLIYATSNTPGVPDSSGDVWKFSTVSGTTRISPPVSDTDTANDFYGY	300
Db	278	RPSHWAGSGDH-WYLSVGDPQPDLMAGGALKWETPAQGRWRWISPIPQ-PASGDGFGW	335
Qy	301	SGLTIDRQHPNTIMVATOISWWPDTIIPRSTDGGAATWRIWDWTSPYNNKSLRYVLDISAE	360
Db	336	GAVAVDPQOPQVLLASTFRRTPRDELYRSVDGKGHWAPL-----LADAVFDHSA	386
Qy	361	PWLTFFGVQPNPVPSPKLGWMDMAIDPNSDRMLYGTGATLYATNDLTKWDSGQIHI	420
Db	387	PW-TAHATPH-----MMG-ALAIIDPFGNHALFVTGYGIWASRNL-----QDFA	428
Qy	421	AP-----MWKGLEETA VNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVF	472
Db	429	APQRLQWFDQGLEETVPLDLLSPMAGAHLLSALGDIIDGFRHDELDEA-QLQYAGPRL	487
Qy	473	TTGTSVDYAEINPSIIVRAGSFDSSSQPNDRHVAFSTDGGKWFQ-GSPFGVTTGGTVA	531
Db	488	TNGESSIDAAQAPQWVRSRGTV-RDRRNEIRALYSRDGGKQMTAFASPPAGQGAGSIA	546
Qy	532	ASADGSRFVWAPGDPGPVVYAVGFGNSWAASQGVPAANAIIRSDRVNPKTFYA--LSNGT	589
Db	547	IGADAAQVWVAPERGGN--WRTSDFGAQQRVDGLPDTAVVMDRVDRRTAVDVASGQ	604
Qy	590	FYRSTDGGVTFPQVAAGL-----PSSGAVGYMFHAVPGEGDLMLAASS	633
Db	605	LYESTDAARSFR--ATGVQVGSPPARDERTRPQLRPDPWRAGVVYLASPGK-----	652
Qy	634	GLYHSTNGSSWSAITGVSSAVNVGFKSAPGSSYPAPVFWCTIGGVTGAYRSDDCGTTW	693
Db	653	GVMRWQDG--TLQVLSQPDSEARSLGKRALRAGAPPALLYLAGRVQGVDFRSDGGVOM	710
Qy	694	VLINDDOHYGNWGOAITGDHANLRVYIGTNGRGIVYGD	733
Db	711	QRINDDAHRFR-PYSVTGDPRIAGRVVYFATGCRGIFYGD	749
RESULT 15			
Q8J0D2	PRELIMINARY; PRT; 812 AA.		
ID	Q8J0D2		
AC	Q8J0D2		
DT	01-MAR-2003	(TrEMBLrel. 23, Created)	
DT	01-MAR-2003	(TrEMBLrel. 23, Last sequence update)	
DT	01-JUN-2003	(TrEMBLrel. 24, Last annotation update)	
DE	Oligoxylolucan reducing end-specific cellobiohydrolase.		
OS	Geotrichum sp. M128		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
OC	Saccharomycetales; Dipodascaceae; mitosporic Dipodascaceae;		
OC	Geotrichum.		
EN	NCBI_TaxID=203496;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=M128;		
RX	MEDLINE=22353052; PubMed=12374797; DOI=10.1074/jbc.M208443200;		
RA	Yaoi K., Mitsuishi Y.;		
RT	"Purification, characterization, cloning, and expression of a novel		

Query Match	28.4%;	Score 1145;	DB 2;	Length 751;
Best Local Similarity	35.0%;	Pred. NO. 3.4e-57;		
Matches 266;	Conservative 121;	Mismatches 295;	Indels 78;	Gaps 21;







GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 07:49:28 ; Search time 28.4312 Seconds  
(without alignments)  
1942.949 Million cell updates/sec

Title: US-09-917-376-3  
Perfect score: 4036  
Sequence: 1 ATQPYTWSNVAIGGGFVD.....YIGTNGRIVYDGGAPSG 740

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues  
Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgm2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgm2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgm2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgm2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgm2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*  
6: /cgm2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	219.5	5.4	2736	US-09-252-991A-30227	Sequence 30227, A
2	181.5	4.5	3892	US-09-328-352-5503	Sequence 5503, Ap
3	167.5	4.2	1164	US-09-902-540-12627	Sequence 12627, A
4	166.5	4.1	1300	US-09-902-540-9932	Sequence 9932, Ap
5	162.5	4.0	772	US-08-410-784A-5	Sequence 5, Appl
6	159.5	4.0	1060	US-09-248-796A-14123	Sequence 14123, A
7	158.5	3.9	774	US-09-346-237-8	Sequence 8, Appl
8	157.5	3.9	688	US-09-489-039A-9813	Sequence 9813, Ap
9	154	3.8	1548	US-09-252-991A-22301	Sequence 22301, A
10	152.5	3.8	1751	US-09-136-574A-44	Sequence 44, Appl
11	152	3.8	3472	US-09-408-020-4	Sequence 4, Appl
12	151.5	3.8	776	US-09-346-237-4	Sequence 4, Appl
13	148.5	3.7	750	5457037-3	Patent No. 5457037
14	148.5	3.7	750	5457037-3	Patent No. 5457037
15	148.5	3.7	751	5457037-5	Patent No. 5457037
16	148.5	3.7	751	5457037-5	Patent No. 5457037
17	148.5	3.7	776	US-09-346-237-7	Sequence 7, Appl
18	147	3.6	3623	US-09-341-461-2	Sequence 2, Appl
19	146	3.6	1187	US-09-949-016-6513	Sequence 6513, Ap
20	146	3.6	1749	US-09-640-419C-28	Sequence 28, Appl
21	146	3.6	2169	US-09-949-016-6930	Sequence 6930, Ap
22	143	3.5	894	US-08-362-525-22	Sequence 22, Appl
23	143	3.5	894	US-08-971-692-15	Sequence 15, Appl
24	143	3.5	4861	US-09-919-497-70	Sequence 70, Appl
25	142	3.5	1216	US-09-134-000C-5130	Sequence 5130, Ap
26	141	3.5	720	US-09-296-284-25	Sequence 25, Appl
27	141	3.5	754	US-09-296-284-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1  
US-09-252-991A-30227  
; Sequence 30227, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 30227  
; LENGTH: 2736  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-30227

Query Match 5.4%; Score 219.5; DB 4; Length 2736;  
Best Local Similarity 21.2%; Pred. No. 6.9e-07;  
Matches 190; Conservative 94; Mismatches 324; Indels 289; Gaps 48;  
QY 59 GNNWGY-----NGV--SIAADPT-NTNKVWA---AVGMYTNSWDNDGAILR-SS 103  
DB 1305 GSGNWTFTPATPLANGTVVNAVAQDPAGNTPGQSGTVDVAVNPVFNVSNGLLNGTA 1364  
QY 104 DQATWQITPLPPKLGNNPGRGMRGLAVDPNNDNILYFCAPSGKGLMRSTDGATWSQ 163  
DB 1365 EPGSTVLT-----DGN--GNPIGQTTA-----DGSNWSFTPG---SQ 1398  
QY 164 MTNFPDVGTYI-ANPTDTTGYQSDIOGVVWVAFDKSSSLGQ---ASKTIFVGVDNPNP 219  
DB 1399 LPN---GTVMVNTASDAAGNTS---APATTTVDSSLPSIPQVDPSPNGSVISGTADAGNT 1451  
QY 220 VF-----WSRDGGATWQAVPCAP-----TGPIR 242  
DB 1452 IIITDGNPNIGQVTDAGSGNWSFTPGIPLPDGTVMNVARSNPNVDSAPAVITVDGVAP 1511  
QY 243 HKGVFPVNVHLYIATNTGGPY---DGSGDVWKFVST-SGTWTRISVPV-----289  
DB 1512 AAVPIDPSNGTEISGTAEAGATVILTDGNGNPGIQAATADGSGNWTFTGPIPLANGTVNA 1571  
QY 290 -STDTANDVFGYGLTIDRQHPNTIMVATOISWMPDITIIFRSTDGGATWTRIDWTSVFN 348  
DB 1572 VAQDPAGNTSGPASVTVDAIAP---PAPVINPSNGVVISGTAEAGAT-VILTDGNGNP- 1625



```
QY 349 RSLRYVLDISAEPWLTFTGVQPNPPV-----PPPKLGWMD-----AMAI 387
Db 1626 -----IGQVTDGSKWAFTPATPLANGTIVINALAQDAAGNSSPSATVDSLAPAAPI 1680
QY 388 DPFNSDRMLYGT---GATLYATNDLTWKDSCGQIHIAPMVKGLBETAVNDLISPPSGAPL 444
Db 1681 DPSNGS-VIAGTAEGAGATVILT-----DNGN-----PIQVQVTDGSGNWSFTT-GRPPL 1727
QY 445 -----ISALG-DLGGFTHADV-TAVPSTIFTPSVF---TTGTSV-DYAEALNPSIIVRAGSF 494
Db 1728 SNGTVVNAVAQDAAGNTSGPASTTVDVAPAPVDPDPSNGSVIAGTAEGATVILTDGG 1787
QY 495 DPSPQNDHRHAFSTDGKNN-----FQSBEPGGVTTGGT----- 529
Db 1788 NPIGA-----TADGSGNSFTPGTPLANGTVINAVAQDPAGNTSGPTSTTVDAVAPA 1840
QY 530 -----VAASADSGSRFWAPGDPQPV--VYVGFNGSNAASGCV-ANAQIRS- 574
Db 1841 TFPVNPNSGVIAGTAEGATVILTDGNGNPIQVQVTDGSGN-WSFTGFTPLANGSVINA 1899
QY 575 -----DRVNPKT-FYALSNGTFYRS-----TDGGVTFQPVAAAG 606
Db 1900 LAQDAAGNTSGPASTTVDVAPAPVDPDPSNGTVISGTAEGATVILTDGG--GNPIRQA 1957
QY 607 LPSSGAVGMFHAFVPGKEGDLWLAASSGLYHSTNG--GSSWSAITGVSSAVNVGFGKAP 664
Db 1958 TADGSGNSFTPGTPLTNGTVINAVAQDAAGNTSGPVSTTVDAVAPFVIDPNSGVKLS 2017
QY 665 GSSVPAPVFWGTIGG-----VTGAYRSDDCGTT----- 692
Db 2018 GTAEPGVRVILTDGNGNPIGQTLADGSGNWTFTPGTPLANGTVVNAVAQDPAGNTSGPAS 2077
QY 693 -----WVLINDQHQYGNWQAITGDHANLRVVGINGRIGVYDGGAPSG 740
Db 2078 TTVDTVAPATPVINASN-----GSVITGT-AEVGAKVILTDGNGNPIGETTADGSG 2127

RESULT 2
US-09-328-352-5503
; Sequence 5503, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5503
; LENGTH: 3892
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5503

Query Match 4.5%; Score 181.5; DB 4; Length 3892;
Best Local Similarity 21.5%; Pred. No. 0.00067;
Matches 204; Conservative 90; Mismatches 344; Indels 309; Gaps 54;

QY 7 TWSNVAIGGGGFVDG-----IVFNEGAPGILYVR-TDIGMYRWDAAANGRWIPLLDW 57
Db 543 TWT-VSVFGSLVADADTKTIDAKVFTDAAGNSSTVNTQI---YTLDTA-APAAPVIDP 597
QY 58 VGNWNWYNGV--VSIADAPINTNKVAAVGMYN-----SWD-PNDGAILRS 102
Db 598 V-----NGTDPITGTAEPGSTVTV-----TYPNGDTATVAGPDGWSVNPNC--LND 643
QY 103 SDQATWQITPLPFKLGN--WPGRMGERLAVDNN-----DNILYFAPSG 148
Db 644 GDEVEAIATDP-----AGNPSLPGTATVD--AVGNTDGVNFTVSDVTADNVINASEASG 696
QY 149 K-----GLNRSTDGATWSQMTNFPDVGTVIANPTDTTGYSQDIQGVVWVAFDKSSSLQG 204
```

```
Db 697 NVTVTGVLKNVPADAANTVTVVINGQVYTTATVDSTAG-----TWTVSVPGSOLTAD 748
QY 205 ASKTI-----FVGADPNPNPVFW 222
Db 749 ADKTIADKVFTDAAGNSVNDTQTYITDITAPDAPVINPVNGTDPITGTAEPSGTVTV 808
QY 223 SRDGATWQAVPGAP-TGFI PHKGVDPVNVHLYIATSNWGPYDGSNGVWKFVSVTSGT 281
Db 809 TYPDGSITTVVAGPDGTWTPNPNGLNDG-DKVTATATDPAGNP-----SLPGT 855
QY 282 WTRISPVSTDTANDYFGYSGLTIDR-----QHPNTIMVATQI 319
Db 856 ATVDVAGNTDGVN--FTVDSVTADNVINASEASGNVTVTVGLKNVPADAANTV-VTVVI 912
QY 320 SWPDTIIFRSTDGATWTRIWDWTSYPNRSIRYVLDISAEPLWTF----- 365
Db 913 NGQVTTATVDSTAG--TWT-----VSVFGSLTADADTKTIDAKVFTDAAGNSVNDTH 965
QY 366 -----GVQNPVPVPSPKLGWMDMAIDPFNSD-----RMLY--GTGATLYATND-- 408
Db 966 TVTVDTVAPNAPV-----LDPINATDPVSGAEPGSTVTVTVPDGTTATVAVGPDGSW 1018
QY 409 -----LTKWDSGQIHI-----APMYKGLSEAVND-----L 435
Db 1019 SVPNPGLNVDGTVTATATADPAGNTSLPGTGVSAADITAPVV-ALDDVLTNDSTPALGT 1077
QY 436 ISPPSGAPLISA-----LGDLGGFTHAD-----VTAVPSTIFTPSVFTTG----- 475
Db 1078 VNDPATTVVNVVDGYDYPVAVNGD-GTWTLANDTLPALTDGPHITVATDAAGNAGTDT 1136
QY 476 --TSVDYAEALNPSIIVRAGSFPDPSQPNDRHVAFTDGGKWNFGQSEPGGVTTGGTVAAS 533
Db 1137 AVVITDITAPNAPVLDPINATDP-----VSGTAEGST-VTVTVPDGTT--ATVWAG 1185
QY 534 ADGSRFWAPGDPQPV-----VYVGFNGSNAASQGVPANAIQSRVNPVKFIYALSNGT 589
Db 1186 TDGS---WSVPNPGNLVDGTVTAT--ATDPAGNTSLPGTGVSAADITAP--VVALDDVL 1238
QY 590 FVRSYTDG--GVTFQFVAAGLPSSGAVGMFHAFVPGKEGDLWLAASSGLYHSTNGSGSWA 647
Db 1239 TNDSTPALGTVNDPTATVVVN--VDGTDYPVAVNGDG-TWTLANDTLPVLADGPHIT-IT 1294
QY 648 ITGVSSAVNVG-----FGKSAPGSSYPAVFVVGTTGGTTCAYRSDDC-----GTTW 693
Db 1295 VTATDAAGNAGTDTAVVITDITAPNA--FVLDPINATDPVSGTAEGSTVTVTVPDGTTA 1352
QY 694 VLI--NDDQHQYGNWQAITGDHANLRVVIQTN-----GRGIVYGD 734
Db 1353 TVVAGTDGWSVNPENGLVDGTVTATATADPAGNTSLPGTGVSAADI 1399

RESULT 3
US-09-902-540-12627
; Sequence 12627, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 12627
; LENGTH: 1164
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-902-540-12627
```







Sequence 5, Application US/08410784A  
Patent No. 5912413  
GENERAL INFORMATION:  
APPLICANT: MYERS, ALAN M.  
APPLICANT: JAMES, MARTHA G.  
TITLE OF INVENTION: ISOLATION OF SUI, A STARCH DEBRANCHING  
ENZYME, THE PRODUCT OF THE MAIZE GENE  
TITLE OF INVENTION: SUGARY 1  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Weingarten, Schurgin, Gagnebin and Hayes LLP  
STREET: Ten Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/410,784A  
FILING DATE: 24-MAR-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Heine, Ph.D., Holliday C.  
REGISTRATION NUMBER: 34,346  
REFERENCE/DOCKET NUMBER: ISU-002XX  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-542-2290  
TELEFAX: 617-451-0313  
TELEX:  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 772 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
US-08-410-784A-5  
Query Match 4.0%; Score 162.5; DB 2; Length 772;  
Best Local Similarity 19.8%; Pred. No. 0.0018;  
Matches 162; Conservative 90; Mismatches 272; Indels 293; Gaps 42;  
QY 51 WPLLDVGNWNNWYN---GVVS---IAADPINTKNWAAVGMVMTNSWDPNDGALLRSSD 104  
Db 108 WGP--NWYASNGKSGAGCVSDVDANGDRPNFKLLDPYQEBSDP----- 155  
QY 105 QGATWQITPLPFKLGNNMGRGMLAVPNDNLYFGAPSGKGLWRSTDSGATWSQM 164  
Db 156 -----LNPSNQ-----GNVFASAHVRTTDSGI----- 178  
QY 165 TNFPDVGYYIANPTDTTG-----YQSDTQGVVWVAFDKSSSLG----- 203  
Db 179 --YAPKGVVLVPSTGTGKTPRAQKDDVIEVHVRG-----FTBQDTSIPAQYRGTYG 231  
QY 204 --QASKTIFGVGA-----DNNPV-----FWSRDGGATWQA 232  
Db 232 AGLKASYLASLGVTAVEFLPVQETQNDANDVVPNSDANQNTWGYNTYFSPDRRYATNK 291  
QY 233 VPGAFTG-FIPKGVFPDVPNVHLYIATNTGTPYDGSNGVWKFVSTGTTWTRISFPVST 291  
Db 292 AAGGTAFQAMQVAFHAGIKVYMDV-----VYNHTAEGGTWSSDPTTAT 338

QY 292 -----DTANDYF-GYSGLTIDRQHNTI---MVATQISWMPDTIIFRSTD 332  
Db 339 IYSMRGLDNATYELTSGNQFYDNTGIGANTYNTVAQNLIVDSVAYWANTM---GVD 395  
QY 333 GGATWTRIDWTSYPNRSLRVLDISAEPWLTFGVQPNPPVSPKLGWDEAMADPPNS 392  
Db 396 GFR-----FDLASVLGNSCLNAVHASA-----PNCPNGGYNFDAADSNVAI----- 436  
QY 393 DRMLY-----GTGATLY-----ATNDLTW-KDSGGOIHIAPMYKGLTEAVNDL 435  
Db 437 NRILREFTRPAAGGTVMICLNGLPSAATR--TSWVDSRRVVRVWVSVPRLRQAOQNEL 494  
QY 436 ISPPSGAPLISALGDLGFTHADVAVPSTIFTSPVFTTGTSDVYAEALNPSIIVR----- 490  
Db 495 -----GSMTIYVTDANDFGSG-----SNLFQSSGRSPWNSINFIDVHDGMTLKDQVISC 543  
QY 491 --AGSFDPSSQPNDRHVAFTDGGKNWFGQSPGGVTTGTV---AASADGSRF-VWAPG 544  
Db 544 NGANNSQASYGFSDG---GTSTNYSWQGSAG---TGAADVQRRRAARTGMAFEMLSAG 596  
QY 545 DP---GQPVVYAVGFN-----SWAASQG---VPANAQIRSDR---VNPX 580  
Db 597 TPLMOGGDEYLRTLQCNNAYNLDSANWLTYSWTTDQSNFTFAQLIRSAKHIFRPS 656  
QY 581 TFYALNNGTFYRSTGVTGTFQFVAAAGLPSSGAVGMFHAVPGKEGDLWLAASGLYHSTN 640  
Db 657 SWYSGSLTWY-----QP-----SGAV-----ADSNYWNNTS 683  
QY 641 GGSSWSAITGVSSAVNVGFGKAPGSSYPAPVFW--GTIGGVTGATRSDDCGTTWVLIND 698  
Db 684 NYAIAVAINGPSL-----GDSNDSIYVAYNGWSSSVTFTLPAPPSTGQMYRVTD 732  
QY 699 DOHQVGNWQAITGDHANLRVYICNGRGIVYGDIG 735  
Db 733 T-----CDWNDGASTFVAPGSETLIG--GAGTTYGCG 763  
RESULT 6  
US-09-248-796A-14123  
Sequence 14123, Application US/09248796A  
Patent No. 6747137  
GENERAL INFORMATION:  
APPLICANT: Keith Weinstock et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.132  
CURRENT APPLICATION NUMBER: US/09/248,796A  
CURRENT FILING DATE: 1999-02-12  
PRIOR APPLICATION NUMBER: US 60/074,725  
PRIOR FILING DATE: 1998-02-13  
PRIOR APPLICATION NUMBER: US 60/096,409  
PRIOR FILING DATE: 1998-08-13  
NUMBER OF SEQ ID NOS: 28208  
SEQ ID NO 14123  
LENGTH: 1060  
TYPE: PXT  
ORGANISM: Candida albicans  
US-09-248-796A-14123  
Query Match 4.0%; Score 159.5; DB 4; Length 1060;  
Best Local Similarity 20.9%; Pred. No. 0.0045;  
Matches 147; Conservative 68; Mismatches 267; Indels 223; Gaps 33;  
QY 7 TWSNVAIGGGGFGVGIWNEGAPGILYVTRDTGG-----MYRWDAANGRW--IPL-LDWV 58  
Db 142 TKYKVPAGYRPPVDAYI-----SATDVNSYTLSYANBYTCAGGYWQRAFFTLRWT 191  
QY 59 GNNV--WGYNGVVSIAADPINTKNWAAVGMVMTNSWDPNDGALLRSSDQATWQI-TPLP 115  
Db 192 GYRNSDAGSNGIVIVATRTTVDSTTAVT---TLPPDPN-----RDKTKTIELKPIP 241  
QY 116 FK-----LGGNMPGRGMLAVPNDNINLYFGAPSGKGLWRST-DSGATW 161



Db 242 TTTTTSYVGVTSYLTAKTAP---IGETATVIV---DIPYHTTTTTSKWTGTTTSTTH 295  
Qy 162 SONTNFPD---VGTIANPT-DTTGYQSDIQGVVWVAEDKSSSSIGQASKTIFVGVADPN 217  
Db 296 TNPTDSIDTVIQVPLPNTVTTTEYSQ---SEATTTITGPGCNTDTVLIRPPP 348  
Qy 218 NPV-----FMSRD---GGATWQAVPCAPTGFPHKGVPDPVHVL-----YIATSNT 261  
Db 349 NHTVTTTEYSSESYTTSTFTAPPGTDSVI---IKEPPNPTVTTTEYSSESYTTTIV 404  
Qy 262 GGPYDG-----SSGDVWKFVSFTGTRISPVSTD-----TAN 295  
Db 405 TAPPGTDTVLIRPPNHTVTTTEYSQSYTT-TTTLVAPPGTDSVIIRPPNPTVTT 463  
Qy 296 DVFYGSGLT-----IDQHPNTMVATQISW-----PDT 325  
Db 464 EYWSQSYATTTTITAPPGETDTVLIRPPNHTVTTTE-YMSQSYATTTTITAPPGETDT 521  
Qy 326 IIFRSTDCGATWTRIDWTSYENRSLRVVLDISABPMLTFGV---QNPVPVPSPKL--- 378  
Db 522 VLIRPPNHTVTT---TEYWSQSYTTTTTTLVAPPGTDSVIIRPPNPTVTTTEYSQ 576  
Qy 379 -----GWMDEAMADPFNSDRMLYGTGATLVATNDLTWKDGGQIHIA----- 421  
Db 577 SVATTTITAPPGTDTVLIRPPNHTVTTSTYWSQSYATTTTITAPAGTDSVIIRPPP 636  
Qy 422 -PMVKGLE-----ETAVNDLISPPSGAPLISALDGLGGTHADTVAVPSTTFTSPVFTTG 475  
Db 637 NPTVTTTEYSQSYTTTNTVTTAPPG-----TDSVIIRPPNPTV 676  
Qy 476 TSDVYAEALNPSIIVRAGSFDPSQPNDRHVAFTDGGKWFQSGPGVT----- 525  
Db 677 TTTEYSQSYATTI-----TVTATPGTDTVLIRPPNPTVTTTTEYSQS 721  
Qy 526 --TGGTVAASADGSRFVWAPGPGQPVVYAVGFGNSWAASQGVPA 568  
Db 722 YATTTVTAAPGDTVLIRPPNPTVTTTTEYSQSYATTTTVA 766

RESULT 7  
US-09-346-237-8  
; Sequence 8, Application US/09346237A  
; Patent No. 6265197  
; GENERAL INFORMATION:  
; APPLICANT: Biesgaard-Frantzen, Henrik  
; APPLICANT: Svendsen, Allan  
; TITLE OF INVENTION: Starch Debranching Enzymes  
; FILE REFERENCE: 5629,200-US  
; CURRENT APPLICATION NUMBER: US/09/346,237A  
; CURRENT FILING DATE: 1999-07-01  
; EARLIER APPLICATION NUMBER: PA 1998 00868  
; EARLIER FILING DATE: 1998-07-02  
; EARLIER APPLICATION NUMBER: 60/094,353  
; EARLIER FILING DATE: 1998-07-28  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 8  
; LENGTH: 774  
; TYPE: PRP  
; ORGANISM: Favobacterium odoratum  
; FEATURE:  
; NAME/KEY: PEPTIDE  
; LOCATION: (1)..(774)  
; OTHER INFORMATION: Isoamylase  
US-09-346-237-8  
Query Match 3.9%; Score 158.5; DB 3; Length 774;  
Best Local Similarity 18.8%; Pred. No. 0.0035;  
Matches 158; Conservative 103; Mismatches 265; Indels 313; Gaps 44;  
Qy 76 INTNKKWAA-----VGMVYNSWPDNGAILRSSDQATWITPL 114  
Db 35 INPNKLGAAYDATKANVTFKWSKATRIELYLYSTGSAEKAKVMTNSGIIWSVT-I 93

Qy 115 PFKLGNPGRMGRLAVDPNNDNILYFPGSGKGLMRSTDSGATWSQMTNPPDVGTYY 174  
Db 94 P---TSTLSGQGLG-----GTLYYGYRANGPNW---PYNASWTGKSSL- 130  
Qy 175 ANPTDTTGYQSDIQG-----VWVAFDKSSSSIGQASKTIFVGVADPNP 219  
Db 131 -----GFISDVDAAGNRFNPNKLSDPYALESHDPTTATM-----TNGS 170  
Qy 220 VWSRDGGATWQAV---PGAPTGF-----PHKGVFDPV---NHLVYIATSN 262  
Db 171 IYAS---GATYRNIDSGSAPKIVLAGDTQATGKPTRALKDDVVYEAHVRGLTWDTS 227  
Qy 263 -----GPYDSSGDVWKFVSFTGTRISPV-----PSTDTANDYFGYSGLT 305  
Db 228 ITAAVRYGTGKAGLAAALGALVTAIEFLPVQETQNDNDPSTSGDNYGYMTLNY 287  
Qy 306 ---DRQ-----HNTIMVATQISWVPTTIIIFRSTDCGATWTRIW 341  
Db 288 PAPDRRYADKTPGPTREFKEMVKAHFDNGIKVLVD-----VYVNHTEGGSW 339  
Qy 342 DMTSPNRSIR-----YVLDISAB-PWLTGCVQNPVPVPS-----PKLGMDEAM 385  
Db 340 DKTTYNTITFRGLDNPYYSLTADFQNSWDNTGVGNYNTRNTIAQNLIVDSLAWRDKL 399  
Qy 386 AIDPFNSDRMLYGTGATLVATNDLTWKDGGQIHIAPMVKGLBEETA VNDLIS--PP--- 439  
Db 400 GVDGVRFDLASVLGNSCQHCFCNFDKWDAG-----NALNRIVAEI LPPRPAT 445  
Qy 440 ---SGAPLIS---ALG---DLGGF-----THADVTAVPSTTFTSPVFTTGTSDV 481  
Db 446 GSGVDLIAEPWALIGNSYQVGGFPGSWAEWNGAYRVDVVRQAQNKLGSAVITTC- 499  
Qy 482 ELNPSIIVR-AGSFDPSQPNDRHVAFTDGGKWFQG---SEPGGVTTGTTVAASADGS 537  
Db 500 ---QMATRFAGSSD-----LYGDDGRKPMHVSFNFITADHGTFLKDLSCNSKN 545  
Qy 538 RFVW---APGDPQPVVYAVGFGNSW-----AASQ-----GVP-----AN 571  
Db 546 NQVMPYGPSDGE-----DNNNSWDQGGIADQKAA RNMALMLLSAGVPMIVGDEA 599  
Qy 572 IRS-----DRVNPKTFY-----ALSNGTFVYRSTD-G 596  
Db 600 LASMNCNNPNYLDSSANLWNSRTTQNNFQSFKAMIAFRKAHPALRPANFYSSVDNN 659  
Qy 597 GVTFPQVAAAGLPSSGAVGVNF-----HAVPGK-EGDLWLAAASGLYHSTNGSSWSAIT 649  
Db 660 GNVMEQLRWFKPDGCVADATYFNDANNHAIAMRIDGSEFGDTASAIYVAHN---AWSA-- 714  
Qy 650 GYSSAVNVGFKSAPGSSYPVAVFVGTGTCVAYRSDDCGTTWVLINDDQHOYGNWGO 708  
Db 715 -----QVNFPLPWPAGKSWYRVTDTGWAEGASQVQAPGSE-ALVGGENTAYGLCLR 766

RESULT 8  
US-09-489-039A-9813  
; Sequence 9813, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709,2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 9813  
; LENGTH: 688  
; TYPE: PRP  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-9813











```

196 DKSSS-----SLGQAKTIFVGVADPNNPVFWS-----RDGATWQAVPG- 235
Db      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
1099 GSGSVHRYSLPSPYAVGAKYETAMIG-GSPGLEFSSDGLRMFVPDAGSETAAVYGL 1157
Qy      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
236 -APTGFIPHKGVDPV-----NHVLYIATSNVTGP----- 264
Db      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
1158 AAPYGI-----GEAPLPPLFLGVGAEBEATLSPDGRHILV-----PGRPGLSQYSLFSTNL 1208
Qy      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
265 -----YDSSGD--VKFVSITSGTWTTRISPVPSDITAN-----DYFGVSG----- 302
Db      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
1209 ELCAEPRIIDGGSCEDGIYAFB-SPGRGEGVSLAASITAADPGIGELHGFAGPMPAPV 1267
Qy      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
303 ---LTIDROH-----PNTIMWAT-----QISWMPDT-----IIFR 329
Db      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
1268 MEQVTLDSRGTLRLVRLDRIVDVTVPYKMWVEDSDGQTTLANSTLLNRENSNILLER 1327
Qy      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
330 STDGA-----TWTRIW---DWTSYNPR-----SLRYLVDLSAEPMLTFGVQ 368
Db      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
1328 LDDAAAGKISYTSVPFRTWSSPLGTGATRPHTLFGDVRLLADIYDASGD----- 1379
Qy      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
369 PNPPVPSPK-LGWDEAMADPNSDRMLYGTGA-----TLYATNDLTKWDSCGQIHA 421
Db      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
1380 -----VPSPSGIEFSDGCRM-----FVTGIGTGGINIFTLSAPFDITLPKHSNSTIG 1428
Qy      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
422 PMVGLBETAVNDLISPPSGAPL-----ISALCD-----LGGPTH-----ADVT-A 461
Db      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
1429 -----GL-----SVSDLAFANGNSLTVLDVGVLRVVALGDDYVVTGTYQKFRITLDTQG 1481
Qy      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
462 VPSTIFTSP-----VFTTGTSDYA-----ELNPSIIVRAGSPDSS-----QP 500
Db      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
1482 IPNSIYTSPDGLSQFVAYDDRIDLVLGSPNDISSTTEIIPYSLPRP---DPPTGMDFTP 1538
Qy      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
501 NDRHVAESTGGKWMFQSGEPGVTG--TVAASADGSRV-----WAGDDP-- 547
Db      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
1539 DGRMFLSTENGIDQYLLSEFAVTTSVFLRTIPIDGGAEGIRFVDRNGRLVFGADGII 1598
Qy      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
548 --QPVVAVGNGSW--AASGVPAQAIRSDRVNPKTFYALSNGTFYRSTGCVTFOPV 603
Db      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
1599 QRHELIIPYGNASTSLLETVRGV-----TDGPGENPA 1631
Qy      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
604 AAGLPSSGAVGVMPHVPKRGDGLWLAASGLYHSTNGSGSSWAITGVSSAVNVGFGKSA 663
Db      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
1632 AGEIRLAGT---FNASDN-----VQSPSGIEFSGDGTGMF---VTGFGAAGVNEFSLSA 1679
Qy      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
664 P-GSSYPANFVVGITGG---VTGAYRSDDCGTTWVLINDD 699
Db      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
1680 PFDTTPLVHVVELHDIGGQPAVDLAFAPED--GRITLLLLAAD 1717

```

RESULT 12

```

US-09-346-237-4
; Sequence 4, Application US/09346237A
; Patent No. 6265197
; GENERAL INFORMATION:
; APPLICANT: Bisgaard-Frantzen, Henrik
; APPLICANT: Svendsen, Allan
; TITLE OF INVENTION: Search Debranching Enzymes
; FILE REFERENCE: 5629.200-US
; CURRENT APPLICATION NUMBER: US/09/346,237A
; CURRENT FILING DATE: 1999-07-01
; EARLIER APPLICATION NUMBER: PA 1998 00868
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: 60/094,353
; EARLIER FILING DATE: 1998-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 776
; TYPE: PRT
; ORGANISM: Pseudomonas amylocleromosa
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(776)

```

```

; OTHER INFORMATION: Isoamylase
US-09-346-237-4
Query Match      3.8%; Score 151.5; DB 3; Length 776;
Best Local Similarity 20.7%; Pred. No. 0.011; Mismatches 291; Gaps 47;
Matches 171; Conservative 89;
Qy      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
27 GAPGILYVRTDIGMYRMDAANGRIPIPLLDWVGNWNGYVVSIAADPINTKNVAAVVG 86
Db      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
96 GITCAVY-----YGYRANGPNWPYASNMGKSGQAGFVSDVDANGDRFNENKLLDPY 147
Qy      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
87 MYTNSWPNDCAILLRSSDQATWQITPLPFLKGMMPGRCGERLAVDNPNNLLYFGAP 146
Db      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
148 AQEVSQDB-----LNPSNQ-----NGNVFASGASYR-----TTDSGIY--AP 182
Qy      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
147 SGKGLMRSTDGATWSQMTNFPDV--GTYIANPTD--TTGYQSDIOGVVWVAFDKSS--SS 201
Db      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
183 KGVVLVSTQSTGTGKPTKRAQKDDVIYEVHVRGFTQDTSIPAQYRGTYYGAGLKASYLAS 242
Qy      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
202 LQQAASKTIFVGVD-----PN---NPVFW-----SRDGGATWQAVPGPTG-F 240
Db      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
243 LG-VTAVEFLPVQETQNDANDVVPNSDANQNYGWTENYFSPDRRYAANKAAGGTAER 301
Qy      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
241 IPHKGVPDPVNHVLYIATSNVTGPGYDSSGDVWKFSTGWTTRISPVST----- 291
Db      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
302 QAMVQAEPHNAGIKVYMDV-----VYNHTAEGGTWTSSDPTTATIYSWRGLDN 348
Qy      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
292 -----DTANDYF-GYSGLTIDROHPNTI---MVATOISWVPTDITIFRSTDCGATWTRIW 341
Db      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
349 ATYYELTSGNQFYDNTGIGANFNNTYTAQNLIVDSLAYWANTM---GVGDGR-----F 400
Qy      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
342 DWTS-YPNRSRLRYLVDLSAEPMLTFGVQPNPVPSPKLGWMDMAID-----PFSNDRML 396
Db      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
401 DLASVLGNSCLNGATASA-----PNCNGGYNFDDAASNVALNRLREFTVRPAA 451
Qy      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
397 YGTGATLYATNDLTKWDSCGQI-H-IAPMVKGLER-----TAVNDLISPPSGAPLI 445
Db      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
452 GSGGLDLFA---EPWALGGSYQLGGFPQGWSEWNLFRDSLQQAQNEL-----GSMTI 502
Qy      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
446 SALDGLGFTHADVTAVPSTIFTSPVFTTGTSDVDAELNPSIIVR-----AGS 493
Db      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
503 YVIQANDFSGS-----SNLFQSGSRSPWNSINFIDVHDGTMTKDVKVSCNGANNQAWP 556
Qy      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
494 FDPSSQPNDRHVARFSTDGK-----NWFGSBEFGVTTGGTV---AASADGSRF-VWAPGD 545
Db      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
557 YGPS-----DGTSTNYSDQCMGAG---TGAADVORRAARTGMAFEMLSAGT 601
Qy      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
546 P-----GQPVVYAVGFGN-----SWAASQGVPAANAQIRSDR-----V 577
Db      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
602 PLMQGGDEYLRTLQCNNAYNLDSSANMLTYSWTTDQ---SNFYTFQRLIAPRKAHPAL 658
Qy      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
578 NPKTFYALSNGTFFYRSTGCVTFQVPAAGLPSSGAVG-----VMFHAVPQKEGD 626
Db      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
659 RPSSWYSGSLTWY-----QP-----SGAVADSNYWNNTSNYAIAYAINGPS-- 700
Qy      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
627 LWLAASSGLYHSTNG-----GSSWSAIT-----GVSSAVNVGFGKSA 665
Db      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
701 --LGDSNSIYVAINGWSSSVTFTLPAPSGTQWYRVITCDWNDDGASTFV-----APG 751
Qy      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
666 SSYPANFVVGITGGVTGAYRSDDCGTTWVLINDDQHQYGNWQAI 710
Db      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
752 SE-----TLIGG-----AGTT-----YQCGGSL 770

```

RESULT 13

```

5457037-3
; Patent No. 5457037
; APPLICANT: TOGNONI, ANGELO;CARRERA, PAOLO;CAMERINI, BARBARA;
; GALLI, GIULIANO;LUCCHESE, GIUSEPPE;GRANDI, GUIDO;DI GENNARO, CARLO
; TITLE OF INVENTION: CLONING OF THE GENE CODING THE ISOAMYLASE
; ENZYME AND ITS USE IN THE PRODUCTION OF SAID ENZYME
; NUMBER OF SEQUENCES: 7
; CURRENT APPLICATION DATA:

```



5457037-3  
; APPLICATION NUMBER: US/08/1,797  
; FILING DATE: 08-JAN-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 749,621  
; FILING DATE: 19-AUG-1991  
; APPLICATION NUMBER: 224,114  
; FILING DATE: 25-JUL-1988  
; SEQ ID NO:3:  
; LENGTH: 750  
5457037-3

Query Match 3.7%; Score 148.5; DB 6; Length 750;  
Best Local Similarity 20.7%; Pred. No. 0.018;  
Matches 171; Conservative 88; Mismatches 275; Indels 291; Gaps 47;  
27 GAPGILYVRTDIGMYRWDAAANGRWIPLLDVWGNMNGYNGVVSIAADPINTNKVWAAVG 86  
70 GITGAVY-----YGYRAGPNWPYASNWKGSQAGFVSDVDANGDRFNPENKLLDPY 121

87 MYTNSWDNDGAILRSSDQATWQITPLPKLGNMPCRGMRGERLAVDPNNDNLLYFGAP 146  
122 AQEVSQDP-----LNPSNQ-----NGNVFASGASYR-----TTDSGIY--AP 156  
147 SGKGLWRSTDGATWSQMTNFPDV--GTYIANPTD--TTGYQSDIQGVVWVAFDKSS--SS 201  
157 KGVLVPSQTGTGKTPRAQKDDVIYEVHVRGFTQDTSIPAYRGTYTYGAGLKASYLAS 216

202 LGQASKTIFVGVAD-----PN-----NPVFW-----SRDGGATWQAVPGAPT-G 240  
217 LG-VTAVEFLVQETQNDANDVVPNSDANQNYGYMTENYFSPDRRYAYNKAAGPTAEF 275  
241 IPHKGVPDPVNHVLYATSNCTGPGYDGSQGVWKFVSVTGWTTRISVPSPST----- 291  
276 QAMVQAFHNAKIKVYMDV-----VYNHTAEGGTWTSDDPTATTIYSWRGLDN 322

292 -----DTANDYF-GYSGLTIDRQHPNTI---MVATQISWMPDITIIFRSTDCGATWTRIW 341  
323 TTYEYELTSGNQFYDNTGIGANFTYNTVAQNLI VDSLAYWANTM---GVDFGR-----F 374  
342 DMTS-YPNRSRLRVLDISAEPLWTFGVQPNPVPSPKLGWMDAEMAD-----PFSNDRML 396  
375 DLASVLGNSCLNGAYTASA-----PNCPNGGYNFDDAADSNAVINRLIEFTVRPAA 425

397 YGTGATLYATNDLTKWDSGGQIH-IAPMVKGLBE-----TAVNDLISPPSGAPLI 445  
426 GSGGLDLFA---EPWALGNSYQLGGFPQGSWSEWNGLFRDSLRLQAQNEL-----GSMITI 476  
446 SALGDLGGFTHADVTAVPSTIFTSVPFTTGTSDVYAEINPSIIVR-----AGS 493  
477 YVTQANDPFGS-----SNLFQSSGRSPWNINFDVHDGMTLKD VVYSCNGANNQAWP 530

494 FDPSSQPNDRHVAFSTDGK-----NWFQSGSEPGVTTGGTV---AASADGSRF-VWAPGD 545  
531 YGPS-----DGGTSTNYSWDQMSAG---TGAANVQRRAARTGMAFEMLSAGT 575  
546 P-----GQPVVYAVGFNG-----SWAASQGVPAQAIRSDR-----V 577  
576 PLMOGDEYLRITLQCNNAYNLDSSANWLTYSWTTDQ---SNFYTFQRLIAFRKAHPAL 632

578 NPKTFYALNGFTFYRSTDCGVTGPVAAGLPSSGAVG-----VMFHAVPCKEGD 626  
633 RPSSWYSGSLTWY-----QP-----SGAVADSNYNWNTSNYAIAYAINGPS-- 674  
627 LWLAASSGLYHSTNG-----GSSWSALT-----GVSSAVNVVFGKSPAG 665  
675 --LGDSNIIYAYNGSSSVTTLFAPPSTGQWYRVTDCTDNDGASTFV-----AEG 725

666 SSYPVAVFVVGITGGVTGYRSDCCGTTWVLINDDCHQYGNWGOAI 710  
726 SE-----TLIGG-----AGTT-----YGCQCGSL 744

5457037-3  
; Patent No. 5457037  
; APPLICANT: TOGNONI, ANGELO; CARRERA, PAOLO; CAMERINI, BARBARA;  
; GALLI, GIULIANO; LUCCHESI, GIUSEPPE; GRANDI, GUIDO; DI GENNARO, CARLO  
; TITLE OF INVENTION: CLONING OF THE GENE CODING THE ISOAMYLASE  
; ENZYME AND ITS USE IN THE PRODUCTION OF SAID ENZYME  
; NUMBER OF SEQUENCES: 7  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/1,797  
; FILING DATE: 08-JAN-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 749,621  
; FILING DATE: 19-AUG-1991  
; APPLICATION NUMBER: 224,114  
; FILING DATE: 25-JUL-1988  
; SEQ ID NO:3:  
; LENGTH: 750  
5457037-3

Query Match 3.7%; Score 148.5; DB 6; Length 750;  
Best Local Similarity 20.7%; Pred. No. 0.018;  
Matches 171; Conservative 88; Mismatches 275; Indels 291; Gaps 47;  
27 GAPGILYVRTDIGMYRWDAAANGRWIPLLDVWGNMNGYNGVVSIAADPINTNKVWAAVG 86  
70 GITGAVY-----YGYRAGPNWPYASNWKGSQAGFVSDVDANGDRFNPENKLLDPY 121

87 MYTNSWDNDGAILRSSDQATWQITPLPKLGNMPCRGMRGERLAVDPNNDNLLYFGAP 146  
122 AQEVSQDP-----LNPSNQ-----NGNVFASGASYR-----TTDSGIY--AP 156  
147 SGKGLWRSTDGATWSQMTNFPDV--GTYIANPTD--TTGYQSDIQGVVWVAFDKSS--SS 201  
157 KGVLVPSQTGTGKTPRAQKDDVIYEVHVRGFTQDTSIPAYRGTYTYGAGLKASYLAS 216

202 LGQASKTIFVGVAD-----PN-----NPVFW-----SRDGGATWQAVPGAPT-G 240  
217 LG-VTAVEFLVQETQNDANDVVPNSDANQNYGYMTENYFSPDRRYAYNKAAGPTAEF 275  
241 IPHKGVPDPVNHVLYATSNCTGPGYDGSQGVWKFVSVTGWTTRISVPSPST----- 291  
276 QAMVQAFHNAKIKVYMDV-----VYNHTAEGGTWTSDDPTATTIYSWRGLDN 322

292 -----DTANDYF-GYSGLTIDRQHPNTI---MVATQISWMPDITIIFRSTDCGATWTRIW 341  
323 TTYEYELTSGNQFYDNTGIGANFTYNTVAQNLI VDSLAYWANTM---GVDFGR-----F 374  
342 DMTS-YPNRSRLRVLDISAEPLWTFGVQPNPVPSPKLGWMDAEMAD-----PFSNDRML 396  
375 DLASVLGNSCLNGAYTASA-----PNCPNGGYNFDDAADSNAVINRLIEFTVRPAA 425

397 YGTGATLYATNDLTKWDSGGQIH-IAPMVKGLBE-----TAVNDLISPPSGAPLI 445  
426 GSGGLDLFA---EPWALGNSYQLGGFPQGSWSEWNGLFRDSLRLQAQNEL-----GSMITI 476  
446 SALGDLGGFTHADVTAVPSTIFTSVPFTTGTSDVYAEINPSIIVR-----AGS 493  
477 YVTQANDPFGS-----SNLFQSSGRSPWNINFDVHDGMTLKD VVYSCNGANNQAWP 530

494 FDPSSQPNDRHVAFSTDGK-----NWFQSGSEPGVTTGGTV---AASADGSRF-VWAPGD 545  
531 YGPS-----DGGTSTNYSWDQMSAG---TGAANVQRRAARTGMAFEMLSAGT 575  
546 P-----GQPVVYAVGFNG-----SWAASQGVPAQAIRSDR-----V 577  
576 PLMOGDEYLRITLQCNNAYNLDSSANWLTYSWTTDQ---SNFYTFQRLIAFRKAHPAL 632

578 NPKTFYALNGFTFYRSTDCGVTGPVAAGLPSSGAVG-----VMFHAVPCKEGD 626  
633 RPSSWYSGSLTWY-----QP-----SGAVADSNYNWNTSNYAIAYAINGPS-- 674  
627 LWLAASSGLYHSTNG-----GSSWSALT-----GVSSAVNVVFGKSPAG 665



Db 675 --LGDSNSIYVAYNGWSSSVTFLLPAPPSTQWYRVTDTCNDWDGASTFV-----APG 725

Qy 666 SSYPANFVVGTTGGVTCAYRSDDCGTTWVLINDQHOYGNWGOAI 710

Db 726 SE-----TLIGG-----AGTT-----YGCCGQSL 744

RESULT 15

5457037-5

Patent No. 5457037

APPLICANT: TOGNONI, ANGELO; CARRERA, PAOLO; CAMERINI, BARBARA;

GALLI, GIULIANO; LUCCHESI, GIUSEPPE; GRANDI, GUIDO; DI GENNARO, CARLO

TITLE OF INVENTION: CLONING OF THE GENE CODING THE ISOMYLASE

ENZYME AND ITS USE IN THE PRODUCTION OF SAID ENZYME

NUMBER OF SEQUENCES: 7

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/1,797

FILING DATE: 08-JAN-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 749,621

FILING DATE: 19-AUG-1991

APPLICATION NUMBER: 224,114

FILING DATE: 25-JUL-1988

SEQ ID NO:5:

LENGTH: 751

5457037-5

Query Match 3.7%; Score 148.5; DB 6; Length 751;

Best Local Similarity 20.7%; Pred. No. 0.018;

Matches 171; Conservative 88; Mismatches 275; Indels 291; Gaps 47;

Qy 27 GAPGILYVRDIDGMYRWDAANGRWIPLLDVWGNNGYGVVSIADPINTNKVMAVG 86

Db 71 GITGAVY-----YGRAWGNWPNYASNWKSGQAGFVSDVDANGDRFNPKLLIDPY 122

Qy 87 MYTNSWDPNDGAILRSSDQATWQITPLPKLGNMPCRMGERLAVDPNNDNILYFCAP 146

Db 123 AQEVSQDP-----LNPSNQ-----NGNVFASGASYR-----TTDSGIY--AP 157

Qy 147 SGKGLMRSTDSGATWSQMTNPPDV--GTYYIANPTD--TTGYQSDIQGVVWVAFDKSS--SS 201

Db 158 KGVVLVFPSTQSTGKTPRAQKDDVIYVHVRGTFEQDTSIPAQYRGTYVYAGLKAASYLAS 217

Qy 202 LQASKTIYFVGAD-----PN---NPVFW-----SRDGGATWQAVGAPTC-F 240

Db 218 LG-VTAVEFLPVQETONDANDWPNSDANQNYGTYMTENYFSPDRRYAYNKAAGGPTAEF 276

Qy 241 IPHKGVPDPVNHVLYIATSNLTGPGPYDGGSGDVWKFVTSGTWTRISVPVST----- 291

Db 277 QAMVQAQHNAGIKYIMDV-----VYNHTAEGGTWTSDDPTATIIYWRGLDN 323

Qy 292 -----DTANDYF--GYSGLTIDRHPNTI---MVATQISWMPDTIIFRSTDGATWTRIW 341

Db 324 TTYVELTSGNQYFYDNTGIGANFNTYNTVAQNLIIVDSLAWNTM---GVDGFR-----F 375

Qy 342 DWTS--YPNRSURYVLDISAEPLWTFGVOPNPVPSPLKGMWDEMAID-----PNSDRML 396

Db 376 DLASVLGNSCLNGAYTASA-----PNCPNGGYNFDAADSNVAINRIIRREFTVRPAA 426

Qy 397 YGTGATLYATNDLTKWDSGGQIH-IAPNVKLEE-----TAVNDLISPPSGAPLI 445

Db 427 GSGGLDLFA-----BFWATGGNSYQLGGPPQGWSEWNGLFRDSLROAQNEL-----GSMTI 477

Qy 446 SALGDLGFTHADVTAVPSTIFTSVPFTTGTSDVYAEINLPSIIVR-----AGS 493

Db 478 YVTQANDFSGS-----SNLFQSSGRSPWNSINFIDVHDGWTLDKDYVSCNGANNSQAWP 531

Qy 494 FDPSSQNDRHVAFSTDGK-----NWFQGSBFGVTTGGTV---AASADGSRF-VWAPGD 545

Db 532 YGPS-----DGGTSTNYSWDQCMGAG---TGAADVQRRAARTGMAFEMLSAGT 576

Qy 546 P-----GQPVVYVAGFN-----SWAASQGVFANAQIRSDR-----V 577

Search completed: October 5, 2005, 08:04:30

Job time : 33.4312 secs

Db 577 PLMQGGDEYLRTLQCNNNAYNLDSANMLTYSWTTDQ-----SNFYTPAQRLIAFRKAHPAL 633

Qy 578 NPKTFYALSNCTFYRSTDDGGVTFFQVAAAGLPSSGAVG-----VMEHAVPGKEGD 636

Db 634 RPSSWYSGQUTWY-----QP-----SGAVADSNYWNNTSNYAIAYAINGPS-- 675

Qy 627 LWLAASSGLYHSTNG-----GSSWSAIT-----GVSSAVNVVFGKSPAG 665

Db 676 --LGDSNSIYVAYNGWSSSVTFLLPAPPSTQWYRVTDTCNDWDGASTFV-----APG 726

Qy 666 SSYPANFVVGTTGGVTCAYRSDDCGTTWVLINDQHOYGNWGOAI 710

Db 727 SE-----TLIGG-----AGTT-----YGCCGQSL 745



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 07:52:23 ; Search time 93.9808 Seconds  
(without alignments)  
3268.602 Million cell updates/sec

Title: US-09-917-376-3  
Perfect score: 4036  
Sequence: 1 ATTQPTWSNVAIGGGGFVD.....YIGTNGRIVGVYDIGGAPSG 740

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1846076 seqs, 415116000 residues

Total number of hits satisfying chosen parameters: 1846076

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA.\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US10E\_PUBCOMB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/1/pubpaa/US11A\_PUBCOMB.pep.\*  
20: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*  
21: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
22: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Length DB ID	Description
1	4036	100.0	740 11	US-09-917-376-3 Sequence 3, Appli
2	4036	100.0	740 11	US-09-917-376-6 Sequence 6, Appli
3	4036	100.0	740 14	US-10-155-400-3 Sequence 3, Appli
4	4036	100.0	740 14	US-10-155-400-6 Sequence 6, Appli
5	4036	100.0	957 11	US-09-917-376-1 Sequence 1, Appli
6	4036	100.0	957 14	US-10-155-400-1 Sequence 1, Appli
7	2429.5	60.2	882 14	US-10-156-761-9395 Sequence 9395, Ap
8	1680	41.6	726 11	US-09-917-376-7 Sequence 7, Appli
9	1680	41.6	726 14	US-10-155-400-7 Sequence 7, Appli
10	1579	39.1	838 15	US-10-420-191-2 Sequence 2, Appli
11	1573.5	39.0	818 14	US-10-026-994-2 Sequence 2, Appli

12	1405	34.8	739	14	US-10-156-761-10111	Sequence 10111, A
13	1117	27.7	789	15	US-10-395-241-14	Sequence 12, Appl
14	1117	27.7	812	15	US-10-395-241-12	Sequence 14, Appl
15	1114	27.6	826	15	US-10-395-241-18	Sequence 18, Appl
16	841.5	20.8	555	10	US-09-927-827-47	Sequence 47, Appl
17	238	5.9	2468	14	US-10-246-330-4	Sequence 4, Appli
18	238	5.9	2468	15	US-10-282-122A-66335	Sequence 66335, A
19	199.5	4.9	2435	15	US-10-282-122A-47453	Sequence 47453, A
20	193	4.8	1465	15	US-10-369-493-13955	Sequence 13955, A
21	181	4.5	1308	15	US-10-282-122A-44566	Sequence 44566, A
22	172.5	4.3	783	15	US-10-369-493-14105	Sequence 14105, A
23	170	4.2	1074	15	US-10-282-122A-50616	Sequence 50616, A
24	168.5	4.2	1439	15	US-10-282-122A-44567	Sequence 44567, A
25	167.5	4.2	1119	14	US-10-245-802-12	Sequence 12, Appl
26	165.5	4.1	613	15	US-10-369-493-11181	Sequence 11181, A
27	163.5	4.1	1289	15	US-10-282-122A-44901	Sequence 44901, A
28	162.5	4.0	772	18	US-10-179-562-39	Sequence 39, Appl
29	158.5	3.9	774	9	US-09-833-435A-8	Sequence 8, Appli
30	158.5	3.9	774	14	US-10-375-720-8	Sequence 8, Appli
31	157	3.9	350	14	US-10-156-761-9225	Sequence 9225, Ap
32	157	3.9	995	16	US-10-437-963-109146	Sequence 109146,
33	156.5	3.9	2204	15	US-10-282-122A-64364	Sequence 64364, A
34	156	3.9	2117	15	US-10-120-801-63	Sequence 63, Appl
35	155	3.8	1096	17	US-10-741-849-7179	Sequence 7179, Ap
36	152	3.8	827	15	US-10-282-122A-47312	Sequence 47312, A
37	152	3.8	1721	15	US-10-027-806-4	Sequence 4, Appli
38	152	3.8	3472	13	US-10-027-806-4	Sequence 4, Appli
39	152	3.8	3472	13	US-10-034-623-4	Sequence 4, Appli
40	152	3.8	3472	14	US-10-027-801-4	Sequence 4, Appli
41	152	3.8	3472	14	US-10-029-120-4	Sequence 4, Appli
42	151.5	3.8	776	9	US-09-833-435A-4	Sequence 4, Appli
43	151.5	3.8	776	14	US-10-375-720-4	Sequence 4, Appli
44	151.5	3.8	2319	17	US-10-831-070-26	Sequence 26, Appl
45	151	3.7	901	17	US-10-939-262-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1  
US-09-917-376-3  
; Sequence 3, Application US/09917376  
; Publication No. US20040038334A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: 40197.4US01  
; CURRENT APPLICATION NUMBER: US/09/917,376  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 740  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; OTHER INFORMATION: Catalytic domain GH74  
US-09-917-376-3

Query Match 100.0%; Score 4036; DB 11; Length 740;				
Best Local Similarity 100.0%; Pred. No. 4.7e-284;				
Matches	740;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
Qy	1	ATTQPTWSNVAIGGGGFVDGIVNEGAPGILYVRTDIGMYRDAANGRIPLDVGW	60	
Db	1	ATTQPTWSNVAIGGGGFVDGIVNEGAPGILYVRTDIGMYRDAANGRIPLDVGW	60	
Qy	61	NNNGYGVSIADPINTNKVAAVGMVYTNWDNDGAILRSSDQGWATQITPLPFLGG	120	



Db 61 NNWNGVSVVSTAADPINTNKVAAVGMVNTSWDPNDGAILRSSDQGTWQITPLPFLKGG 120  
QY 121 NMPGRGMRGLAVDPNNNDNIIYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDT 180  
Db 121 NMPGRGMRGLAVDPNNNDNIIYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDT 180  
QY 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIIFVGADPNPNPVFWSRGGATWQAVPGAPTGF 240  
Db 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIIFVGADPNPNPVFWSRGGATWQAVPGAPTGF 240  
QY 241 IPHKGVFPVNVHLYIATSNTPGPGYDSSGDVWKFVSVTSGTWTRISPVPSSTDANDYFGY 300  
Db 241 IPHKGVFPVNVHLYIATSNTPGPGYDSSGDVWKFVSVTSGTWTRISPVPSSTDANDYFGY 300  
QY 301 SGLTIDRQHPNTIMVATOISWMPDTIIFRSTDDGATWTRIMDWTSPNRSRLRYVLDISAE 360  
Db 301 SGLTIDRQHPNTIMVATOISWMPDTIIFRSTDDGATWTRIMDWTSPNRSRLRYVLDISAE 360  
QY 361 PMLTFGVQPNPVPSPKLGWDEAMAIIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420  
Db 361 PMLTFGVQPNPVPSPKLGWDEAMAIIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420  
QY 421 APWKGLBEETAVNDLISPPSGAPLISALDGLGFFHADVTAVPSTIPTSPTTGTSTVDY 480  
Db 421 APWKGLBEETAVNDLISPPSGAPLISALDGLGFFHADVTAVPSTIPTSPTTGTSTVDY 480  
QY 481 AELNPSIIVRAGSPDPSSQPNDRHVAFSTDDGKNWFOGSEPGVTTGGTVAASADGSRFV 540  
Db 481 AELNPSIIVRAGSPDPSSQPNDRHVAFSTDDGKNWFOGSEPGVTTGGTVAASADGSRFV 540  
QY 541 WAPDGPQPVVYAVGFGNSWAASQGVAPANAQIRSDRVNPKTFYALNSNGTFYRSTDDGGVTF 600  
Db 541 WAPDGPQPVVYAVGFGNSWAASQGVAPANAQIRSDRVNPKTFYALNSNGTFYRSTDDGGVTF 600  
QY 601 QPVAAGLPSSGAVGVWFMHAFVPGKGDMLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 660  
Db 601 QPVAAGLPSSGAVGVWFMHAFVPGKGDMLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 660  
QY 661 KSAPGSSYPVAVFVVTIGVTCVAYRSDCGTTWVLINDDQHOYGNWGOAITGDHANLRRV 720  
Db 661 KSAPGSSYPVAVFVVTIGVTCVAYRSDCGTTWVLINDDQHOYGNWGOAITGDHANLRRV 720  
QY 721 YIGTNGRGIVYGDIGGAPSG 740  
Db 721 YIGTNGRGIVYGDIGGAPSG 740

## RESULT 2

US-09-917-376-6  
; Sequence 6, Application US/09917376  
; Publication No. US20040038334A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
; FILE REFERENCE: 40197.4US01  
; CURRENT APPLICATION NUMBER: US/09/917,376  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 740  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
US-09-917-376-6

Query Match 100.0%; Score 4036; DB 11; Length 740;  
Best Local Similarity 100.0%; Pred. No. 4.7e-284;  
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTQPTYSNVAIGGGFVGDGIVFNEGAPGILYVTRTDIGCMYRWDAANGRWIPLLDWVGW 60  
Db 1 ATTQPTYSNVAIGGGFVGDGIVFNEGAPGILYVTRTDIGCMYRWDAANGRWIPLLDWVGW 60  
QY 61 NNWNGVSVVSTAADPINTNKVAAVGMVNTSWDPNDGAILRSSDQGTWQITPLPFLKGG 120  
Db 61 NNWNGVSVVSTAADPINTNKVAAVGMVNTSWDPNDGAILRSSDQGTWQITPLPFLKGG 120  
QY 121 NMPGRGMRGLAVDPNNNDNIIYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDT 180  
Db 121 NMPGRGMRGLAVDPNNNDNIIYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDT 180  
QY 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIIFVGADPNPNPVFWSRGGATWQAVPGAPTGF 240  
Db 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIIFVGADPNPNPVFWSRGGATWQAVPGAPTGF 240  
QY 241 IPHKGVFPVNVHLYIATSNTPGPGYDSSGDVWKFVSVTSGTWTRISPVPSSTDANDYFGY 300  
Db 241 IPHKGVFPVNVHLYIATSNTPGPGYDSSGDVWKFVSVTSGTWTRISPVPSSTDANDYFGY 300  
QY 301 SGLTIDRQHPNTIMVATOISWMPDTIIFRSTDDGATWTRIMDWTSPNRSRLRYVLDISAE 360  
Db 301 SGLTIDRQHPNTIMVATOISWMPDTIIFRSTDDGATWTRIMDWTSPNRSRLRYVLDISAE 360  
QY 361 PMLTFGVQPNPVPSPKLGWDEAMAIIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420  
Db 361 PMLTFGVQPNPVPSPKLGWDEAMAIIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420  
QY 421 APWKGLBEETAVNDLISPPSGAPLISALDGLGFFHADVTAVPSTIPTSPTTGTSTVDY 480  
Db 421 APWKGLBEETAVNDLISPPSGAPLISALDGLGFFHADVTAVPSTIPTSPTTGTSTVDY 480  
QY 481 AELNPSIIVRAGSPDPSSQPNDRHVAFSTDDGKNWFOGSEPGVTTGGTVAASADGSRFV 540  
Db 481 AELNPSIIVRAGSPDPSSQPNDRHVAFSTDDGKNWFOGSEPGVTTGGTVAASADGSRFV 540  
QY 541 WAPDGPQPVVYAVGFGNSWAASQGVAPANAQIRSDRVNPKTFYALNSNGTFYRSTDDGGVTF 600  
Db 541 WAPDGPQPVVYAVGFGNSWAASQGVAPANAQIRSDRVNPKTFYALNSNGTFYRSTDDGGVTF 600  
QY 601 QPVAAGLPSSGAVGVWFMHAFVPGKGDMLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 660  
Db 601 QPVAAGLPSSGAVGVWFMHAFVPGKGDMLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 660  
QY 661 KSAPGSSYPVAVFVVTIGVTCVAYRSDCGTTWVLINDDQHOYGNWGOAITGDHANLRRV 720  
Db 661 KSAPGSSYPVAVFVVTIGVTCVAYRSDCGTTWVLINDDQHOYGNWGOAITGDHANLRRV 720  
QY 721 YIGTNGRGIVYGDIGGAPSG 740  
Db 721 YIGTNGRGIVYGDIGGAPSG 740

## RESULT 3

US-10-155-400-3  
; Sequence 3, Application US/10155400  
; Publication No. US20030108988A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
; FILE REFERENCE: NREL 01-36A  
; CURRENT APPLICATION NUMBER: US/10/155,400  
; CURRENT FILING DATE: 2002-10-22  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 740  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus



FEATURE:  
; OTHER INFORMATION: Catalytic domain GH74  
US-10-155-400-3

Query Match 100.0%; Score 4036; DB 14; Length 740;  
Best Local Similarity 100.0%; Pred. No. 4.7e-284;  
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTQPTVTSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDANGRWIPLLDWVGW 60  
DB 1 ATTQPTVTSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDANGRWIPLLDWVGW 60  
QY 61 NNWGYGVVSTAADPINTNKVAAVGMVYNSWDNDGAILRSSDQATWQITPLPFLKGG 120  
DB 61 NNWGYGVVSTAADPINTNKVAAVGMVYNSWDNDGAILRSSDQATWQITPLPFLKGG 120  
QY 121 NMPGRGMGERLAVDPNNNDNILYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDT 180  
DB 121 NMPGRGMGERLAVDPNNNDNILYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDT 180  
QY 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIIFVGADPNPFWRSRGGATWQAVPGAPTGF 240  
DB 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIIFVGADPNPFWRSRGGATWQAVPGAPTGF 240  
QY 241 IPHKGVDPNVNHVLIATSNTPGPGYDSSGVWKFVSCTGTWTRISPVSTDTANDYFGY 300  
DB 241 IPHKGVDPNVNHVLIATSNTPGPGYDSSGVWKFVSCTGTWTRISPVSTDTANDYFGY 300  
QY 301 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGATWTRIIDWTSYPNRSRLRYVLDISAE 360  
DB 301 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGATWTRIIDWTSYPNRSRLRYVLDISAE 360  
QY 361 PWTFTGVQPNPVPSPKLGWMDMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420  
DB 361 PWTFTGVQPNPVPSPKLGWMDMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420  
QY 421 APWKGLEBETAVNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTGSVDY 480  
DB 421 APWKGLEBETAVNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTGSVDY 480  
QY 481 AELNPSIIVRAGSFPDPSQPNDRHVAFTDGGKNWFGSEPGVTTGGTVAASADGSRFV 540  
DB 481 AELNPSIIVRAGSFPDPSQPNDRHVAFTDGGKNWFGSEPGVTTGGTVAASADGSRFV 540  
QY 541 WAPDGPQPVVYAVGFGNSWAASQGVPAQAQIRSDRVNPKTFYALSNGTFYRSTDGGVTF 600  
DB 541 WAPDGPQPVVYAVGFGNSWAASQGVPAQAQIRSDRVNPKTFYALSNGTFYRSTDGGVTF 600  
QY 601 QPVAAGLPSSGAVGVNMFHAPVKGEDLMLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660  
DB 601 QPVAAGLPSSGAVGVNMFHAPVKGEDLMLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660  
QY 661 KSAPGSSYPVAVFVGTIGVTVGAYRSDCGTTWVLINDDQHOYGNWQAITGDHANLRV 720  
DB 661 KSAPGSSYPVAVFVGTIGVTVGAYRSDCGTTWVLINDDQHOYGNWQAITGDHANLRV 720  
QY 721 YIGTNGRGIVYGDIGGAPSG 740  
DB 721 YIGTNGRGIVYGDIGGAPSG 740

RESULT 4  
US-10-155-400-6  
; Sequence 6, Application US/10155400  
; Publication No. US20030108988A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
; FILE REFERENCE: NREL 01-36A

CURRENT APPLICATION NUMBER: US/10/155,400  
; CURRENT FILING DATE: 2002-10-22  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 740  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulosilyticus  
US-10-155-400-6

Query Match 100.0%; Score 4036; DB 14; Length 740;  
Best Local Similarity 100.0%; Pred. No. 4.7e-284;  
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTQPTVTSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDANGRWIPLLDWVGW 60  
DB 1 ATTQPTVTSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDANGRWIPLLDWVGW 60  
QY 61 NNWGYGVVSTAADPINTNKVAAVGMVYNSWDNDGAILRSSDQATWQITPLPFLKGG 120  
DB 61 NNWGYGVVSTAADPINTNKVAAVGMVYNSWDNDGAILRSSDQATWQITPLPFLKGG 120  
QY 121 NMPGRGMGERLAVDPNNNDNILYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDT 180  
DB 121 NMPGRGMGERLAVDPNNNDNILYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDT 180  
QY 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIIFVGADPNPFWRSRGGATWQAVPGAPTGF 240  
DB 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIIFVGADPNPFWRSRGGATWQAVPGAPTGF 240  
QY 241 IPHKGVDPNVNHVLIATSNTPGPGYDSSGVWKFVSCTGTWTRISPVSTDTANDYFGY 300  
DB 241 IPHKGVDPNVNHVLIATSNTPGPGYDSSGVWKFVSCTGTWTRISPVSTDTANDYFGY 300  
QY 301 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGATWTRIIDWTSYPNRSRLRYVLDISAE 360  
DB 301 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGATWTRIIDWTSYPNRSRLRYVLDISAE 360  
QY 361 PWTFTGVQPNPVPSPKLGWMDMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420  
DB 361 PWTFTGVQPNPVPSPKLGWMDMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420  
QY 421 APWKGLEBETAVNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTGSVDY 480  
DB 421 APWKGLEBETAVNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTGSVDY 480  
QY 481 AELNPSIIVRAGSFPDPSQPNDRHVAFTDGGKNWFGSEPGVTTGGTVAASADGSRFV 540  
DB 481 AELNPSIIVRAGSFPDPSQPNDRHVAFTDGGKNWFGSEPGVTTGGTVAASADGSRFV 540  
QY 541 WAPDGPQPVVYAVGFGNSWAASQGVPAQAQIRSDRVNPKTFYALSNGTFYRSTDGGVTF 600  
DB 541 WAPDGPQPVVYAVGFGNSWAASQGVPAQAQIRSDRVNPKTFYALSNGTFYRSTDGGVTF 600  
QY 601 QPVAAGLPSSGAVGVNMFHAPVKGEDLMLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660  
DB 601 QPVAAGLPSSGAVGVNMFHAPVKGEDLMLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660  
QY 661 KSAPGSSYPVAVFVGTIGVTVGAYRSDCGTTWVLINDDQHOYGNWQAITGDHANLRV 720  
DB 661 KSAPGSSYPVAVFVGTIGVTVGAYRSDCGTTWVLINDDQHOYGNWQAITGDHANLRV 720  
QY 721 YIGTNGRGIVYGDIGGAPSG 740  
DB 721 YIGTNGRGIVYGDIGGAPSG 740

RESULT 5  
US-09-917-376-1  
; Sequence 1, Application US/09917376  
; Publication No. US20040038334A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU



APPLICANT: ADNEY, WILLIAM S.  
APPLICANT: VINZANT, TODD B.  
APPLICANT: HIMMEL, MICHAEL E.  
TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
TITLE OF INVENTION: CELLULOYTICUS  
FILE REFERENCE: 40197.4US01  
CURRENT APPLICATION NUMBER: US/09/917.376  
CURRENT FILING DATE: 2001-07-28  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 1  
LENGTH: 957  
TYPE: PRT  
ORGANISM: Acidothermus cellulolyticus  
FEATURE:  
NAME/KEY: MOD RES  
LOCATION: (957)  
OTHER INFORMATION: Any amino acid  
US-09-917-376-1

Query Match 100.0%; Score 4036; DB 11; Length 957;  
Best Local Similarity 100.0%; Pred. No. 6.6e-284;  
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTQPYTMSNVAIGGGFVDGIVFNEGAPGILYVRTDIGNMYRWDAAANGRWIPLLDWVGW 60  
Db 47 ATTQPYTMSNVAIGGGFVDGIVFNEGAPGILYVRTDIGNMYRWDAAANGRWIPLLDWVGW 106  
Qy 61 NNWYNGVVSAAADPINTNKVAAVGMVNTSWDNDGAILRSSDQATWQITPLPFKLG 120  
Db 107 NNWYNGVVSAAADPINTNKVAAVGMVNTSWDNDGAILRSSDQATWQITPLPFKLG 166  
Qy 121 NMPGRGGERLAVDPNNDNIIYFGAPSGKGLWRSTDGATWSQMTNPDVGYIANPTDT 180  
Db 167 NMPGRGGERLAVDPNNDNIIYFGAPSGKGLWRSTDGATWSQMTNPDVGYIANPTDT 226  
Qy 181 TGYQSDIQGVVWVAFDKSSSSLGQASKTIFVGVADPNPNVFWSDGGATWQAVPGAPTGF 240  
Db 227 TGYQSDIQGVVWVAFDKSSSSLGQASKTIFVGVADPNPNVFWSDGGATWQAVPGAPTGF 286  
Qy 241 IPHKGVPDPVNHVLIATSNTPGPGYDSSGDVWKFVSVTSGTWTRISPVPSDTANDYFGY 300  
Db 287 IPHKGVPDPVNHVLIATSNTPGPGYDSSGDVWKFVSVTSGTWTRISPVPSDTANDYFGY 346  
Qy 301 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGATWTRIWDWTSYPNRSRYVLDISAE 360  
Db 347 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGATWTRIWDWTSYPNRSRYVLDISAE 406  
Qy 421 APWKGLLEETAANDLISPPSGAPLISALGDLGGFTHADVAVPSTIPTSVPFTTGSVDY 480  
Db 467 APWKGLLEETAANDLISPPSGAPLISALGDLGGFTHADVAVPSTIPTSVPFTTGSVDY 526  
Qy 481 AELNPSIIVRAGSDPSSQPNDRHVAFTDGGKWFQSGEPGGVTTGGTVAASADGSRFV 540  
Db 527 AELNPSIIVRAGSDPSSQPNDRHVAFTDGGKWFQSGEPGGVTTGGTVAASADGSRFV 586  
Qy 541 WAPGDGQPVVYVGVFGNSWAAASQGVPAQAQIRSDRVNPKTFYALSNGTFYRSTDGGVTF 600  
Db 587 WAPGDGQPVVYVGVFGNSWAAASQGVPAQAQIRSDRVNPKTFYALSNGTFYRSTDGGVTF 646  
Qy 601 QPVAAGLPSSGAVGMFHVPGKSGDLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 660  
Db 647 QPVAAGLPSSGAVGMFHVPGKSGDLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 706  
Qy 661 KSAPGSSYPVAVVGTIGVTVGAYRSDDCGTTWVLINDDQHQYGNWGOAITGDHANLRRV 720  
Db 707 KSAPGSSYPVAVVGTIGVTVGAYRSDDCGTTWVLINDDQHQYGNWGOAITGDHANLRRV 766  
Qy 721 YIGTNGRGIVYDGGAPSG 740

Db 767 YIGTNGRGIVYDGGAPSG 786

RESULT 6  
US-10-155-400-1  
Sequence 1, Application US/10155400  
Publication No. US20030108988A1  
GENERAL INFORMATION:  
APPLICANT: DING, SHI-YOU  
APPLICANT: ADNEY, WILLIAM S.  
APPLICANT: VINZANT, TODD B.  
APPLICANT: HIMMEL, MICHAEL E.  
TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
TITLE OF INVENTION: CELLULOYTICUS  
FILE REFERENCE: NREL 01-36A  
CURRENT APPLICATION NUMBER: US/10/155,400  
CURRENT FILING DATE: 2002-10-22  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 1  
LENGTH: 957  
TYPE: PRT  
ORGANISM: Acidothermus cellulolyticus  
FEATURE:  
NAME/KEY: MOD RES  
LOCATION: (957)  
OTHER INFORMATION: Any amino acid  
US-10-155-400-1

Query Match 100.0%; Score 4036; DB 14; Length 957;  
Best Local Similarity 100.0%; Pred. No. 6.6e-284;  
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTQPYTMSNVAIGGGFVDGIVFNEGAPGILYVRTDIGNMYRWDAAANGRWIPLLDWVGW 60  
Db 47 ATTQPYTMSNVAIGGGFVDGIVFNEGAPGILYVRTDIGNMYRWDAAANGRWIPLLDWVGW 106  
Qy 61 NNWYNGVVSAAADPINTNKVAAVGMVNTSWDNDGAILRSSDQATWQITPLPFKLG 120  
Db 107 NNWYNGVVSAAADPINTNKVAAVGMVNTSWDNDGAILRSSDQATWQITPLPFKLG 166  
Qy 121 NMPGRGGERLAVDPNNDNIIYFGAPSGKGLWRSTDGATWSQMTNPDVGYIANPTDT 180  
Db 167 NMPGRGGERLAVDPNNDNIIYFGAPSGKGLWRSTDGATWSQMTNPDVGYIANPTDT 226  
Qy 181 TGYQSDIQGVVWVAFDKSSSSLGQASKTIFVGVADPNPNVFWSDGGATWQAVPGAPTGF 240  
Db 227 TGYQSDIQGVVWVAFDKSSSSLGQASKTIFVGVADPNPNVFWSDGGATWQAVPGAPTGF 286  
Qy 241 IPHKGVPDPVNHVLIATSNTPGPGYDSSGDVWKFVSVTSGTWTRISPVPSDTANDYFGY 300  
Db 287 IPHKGVPDPVNHVLIATSNTPGPGYDSSGDVWKFVSVTSGTWTRISPVPSDTANDYFGY 346  
Qy 301 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGATWTRIWDWTSYPNRSRYVLDISAE 360  
Db 347 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGATWTRIWDWTSYPNRSRYVLDISAE 406  
Qy 361 PWLTFGVQPNPVPSPKLGWMDAMAIDPFNSDRMLYGTGATLYATNDLTWDSGGQIHI 420  
Db 407 PWLTFGVQPNPVPSPKLGWMDAMAIDPFNSDRMLYGTGATLYATNDLTWDSGGQIHI 466  
Qy 421 APWKGLLEETAANDLISPPSGAPLISALGDLGGFTHADVAVPSTIPTSVPFTTGSVDY 480  
Db 467 APWKGLLEETAANDLISPPSGAPLISALGDLGGFTHADVAVPSTIPTSVPFTTGSVDY 526  
Qy 481 AELNPSIIVRAGSDPSSQPNDRHVAFTDGGKWFQSGEPGGVTTGGTVAASADGSRFV 540  
Db 527 AELNPSIIVRAGSDPSSQPNDRHVAFTDGGKWFQSGEPGGVTTGGTVAASADGSRFV 586  
Qy 541 WAPGDGQPVVYVGVFGNSWAAASQGVPAQAQIRSDRVNPKTFYALSNGTFYRSTDGGVTF 600  
Db 587 WAPGDGQPVVYVGVFGNSWAAASQGVPAQAQIRSDRVNPKTFYALSNGTFYRSTDGGVTF 646



QY 426 GLEETA VNDLISPPS-GAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTGISVDYAE LN 484  
Db 456 GLEETA VNDLASPPSGAQLFSALGDIIGFRHTDLTTPSLMYTSPNFTTSTSLDYAETD 515  
QY 485 PSIIYRAGSFDPSOPNDRHVAFSTDGKNWFQSGEPGGVTTGGTVAASADGSRFVWAPG 544  
Db 516 PGTVVVRVGNLD--SGP---HVAFSTDNGANWFAGADPSVGGGTVAASADGSRFVWSPA 570  
QY 545 DPGQPWYAVGFGNSWAASQGVPAQAQRSDRVNPKRTFYALNSNGTFFRSTGDTGVPQVA 604  
Db 571 GTG--VQYTTGFGTWSASAGLPAGAIVESDRVDPKTFYFGKSGRFFVSSDGGATPTASA 628  
QY 605 A-GLPSSGAVGYMFHAVPGKEGDLMAASS-----GLYHSTNGGSSWSALTGVSSAVNVG 658  
Db 629 ATGLPSGDS--VRFKALPGTKGDIWLAGGASGAYGLMHSITDGAFAFTKLATVDQADTIG 686  
QY 659 FOKSAPGSSYPAVFVGTIGTGYRSDDCGTTWVLINDDOHOYGNWGOAITGDHANLR 718  
Db 687 FGKAATGASYQTLTYSAKIGGVGRGIFRSTDKGASWTRVNDDAHQWGTGAITGDPRVYG 746  
QY 719 RVYGTNGRGIVYDGGAPSG 740  
Db 747 RVYVSTNGRGIVGTAGSSDG 768  
RESULT 8  
US-09-917-376-7  
; Sequence 7, Application US/09917376  
; Publication No. US2004003834A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOYTICUS  
; FILE REFERENCE: 40197.4US01  
; CURRENT APPLICATION NUMBER: US/09/917,376  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 7  
; LENGTH: 726  
; TYPE: PRT  
; ORGANISM: Aspergillus aculeatus  
US-09-917-376-7  
Query Match 41.6%; Score 1680; DB 11; Length 726;  
Best Local Similarity 46.2%; Pred. No. 4.7e-113;  
Matches 346; Conservative 113; Mismatches 258; Indels 32; Gaps 17;  
QY 1 ATTQPTWNSVAI-GGGGFVDGIVFNEGAPGILYVTRDIGMYRWDAAANGRIWILLDDWVG 59  
Db 1 AASQAYTWKVVVTCGGGFTPGIVFNPSPAKGVAYARTDIGAYRLN--SDDTWTPLMDWVG 59  
QY 60 ---WNNWYNGVVSIAADPINTKVAACVMTNSWDNDGATILRSSDQCATQITPLPF 116  
Db 60 NDTWHDW---GIDALATDPVDTDRVYVAVGMTNWDPNVGSILRSTDDGTWTETKLPF 116  
QY 117 KLGNNPGRGMGRRLAVDPNNDNILYFGAPSGKGLNRSTDGATWSQMTNPPDVGTYIAN 176  
Db 117 KVGNPMPGRGMGRRLAVDPNKNILYFGARSGHGLWKSTDYGATWSNVTSFTWTGYTQD 176  
QY 177 PTDITGYSQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNNPVFWSRDGGATQWAVPGA 236  
Db 177 SSST--YTSDPVGIAMVTFDSTSGSGSATPRIFVGVADAGKSVFKSEDACATWAWVSGE 234  
QY 237 PT-GFTPHKGVDPVNHVLIATSNCTGPGYDGSSEDVWKFSTVSGTWTIRISVPSTDTAN 295  
Db 235 PQYGFPHKGVLSPEEKTLYISYANGAGPYDGTNGTGHKYNITSVGVWTDISP---TSLAS 291  
QY 296 DYFGYSLGTLTDROHPNTIMVATQISWMPDITIFRSTDDGATWTRIMDWTYSYPNRSLRYVL 355

QY 601 QPVAAGLPSSGAVGYMFHAVPGKEGDLWLAASSGLYHSTNGGSSWSALTGVSSAVNVGFG 660  
Db 647 QPVAAGLPSSGAVGYMFHAVPGKEGDLWLAASSGLYHSTNGGSSWSALTGVSSAVNVGFG 706  
QY 661 KSAPOSSYPVAVFVGTIGTGYRSDDCGTTWVLINDDOHOYGNWGOAITGDHANLRV 720  
Db 707 KSAPOSSYPVAVFVGTIGTGYRSDDCGTTWVLINDDOHOYGNWGOAITGDHANLRV 766  
QY 721 YIGTNGRGIVYDGGAPSG 740  
Db 767 YIGTNGRGIVYDGGAPSG 786  
RESULT 7  
US-10-156-761-9395  
; Sequence 9395, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 9395  
; LENGTH: 882  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-9395  
Query Match 60.2%; Score 2429.5; DB 14; Length 882;  
Best Local Similarity 59.0%; Pred. No. 2.4e-167;  
Matches 438; Conservative 111; Mismatches 174; Indels 19; Gaps 8;  
QY 6 YTSNVAIGGGFVDGIVFNEGAPGILYVTRDIGMYRWDAAANGRIWILLDDWVGNWNGY 65  
Db 39 YSWKARVDDGGFVPGIVFNRSEKRLAYARTDIGMYRWAESSKTWTPLLDSVGSWDGCH 98  
QY 66 NGVVSIAADPINTKVAACVMTNSWDNDGATILRSSDQCATQITPLPFKLGNNMPGR 125  
Db 99 TGVSILASDSDPNKYAAVGTNTNSWDPCGAVLRSGBRGASWQKTDLPFKLGNNMPGR 158  
QY 126 GNGERLAVDPNNDNILYFGAPSGKGLNRSTDGATWSQMTNPPDVGTYIANPDTTGYOS 185  
Db 159 GNGERLAVDPNNSVLYLGPAGSKGLNRSTDGATWSQMTNPPDVGTYIANPDTTGYOS 218  
QY 186 DIQGVVWVAFDKSSSLGQASKTIFVGVADPNNPVFWSRDGGATQWAVCAPGTIPHKG 245  
Db 219 DNQIGVWVTFDSTGPGSGSTRIVYGVADKONSVYRSTDTAGATWSRLAQPTGHLAHKG 278  
QY 246 VFDPVNHVLIATSNCTGPGYDGSSEDVWKFSTVSGTWTIRISVPSTDTANDYFGYSLGTI 305  
Db 279 VLDAANGCLYLAYSVDKGGYDGGKQLWRYTTKTGTWNTISPVAEADT---YVFGSLTIV 335  
QY 306 DRQHPNTIMVATQISWMPDITIFRSTDDGATWTRIMDWTYSYPNRSLRYVLDTISAEWLPF 365  
Db 336 DRQHPCTVMATAYSSWMPDITQFRSTDSGTWTKAWDYTSYPSRSNRFMTDVSSESPWLTW 395  
QY 366 GVQPNPVPSPKLGWDEAWADPNSDRMLYGTGATLYATNDLTKWDSGGQIHIAPMVK 425  
Db 396 GANPAPPEQTPLKGMWTESLEIDPFDSARMYGTGATVYGTDLNLTWDSGSGQTIKPMAR 455



```
Db 292 TTYGGLSVLDQVPGTLMVAALNCWPDDELIFRSTDSGATWSPIEWNGVPSINYYYS 351
Qy 356 DISAEPMLTFGVQNP-PVPSPKLGMDEMAIDPFNSDRMLYGTGATLYATNDLTWDS 414
Db 352 DISAPWLODTTSDQFPV--RVGMVEALAIIDPFDSNHWLYGTGLTVYGGHDLTNWDS 408
Qy 415 GGQIHIAPMVKLETA VNDLISPPSGAPLISALGDLGGFTHADVAVPSTIFTSPVFTT 474
Db 409 KHNVTVKSLAVGIEEMAVLGLITPPGPGALLSAVGDDGCFYHSDLDAAAPNQAHTPTTGT 468
Qy 475 GTSVDYAEALNPSIIVRAGSPSPQPNDRHVAFTSDGKNWFGQSEPGVTTGGTVAASA 534
Db 469 TNGIDYAGNKPNSNIVRSGASD--DYPT--LALSNFGSTWYADYAASTSTGTGAVALSA 523
Qy 535 DGSRFVWAPGDPGQPVVAVGFGNSWAASQGVPAQAIRSDRVPNPKTFYALSNGTFYRST 594
Db 524 DGDVTLMSSTSGALVSKSQG---TLTAVSSLPSCGAVIASDKSDNTVFGGSAGAIYVSK 580
Qy 595 DGGVTFQVAAAGLPSGAGVGMFHAVPGKEGDLMLAASSGLYHSTNGGSSWSAI-TGVSS 653
Db 581 NTATSFTKTVS-LGSSSTTVNAI-RAHPSIAGDVMASTDGKLWHSTDYGSTTQIGSGVTA 638
Qy 654 AVNVFGKSPGSSYPAPVVGTTGGVTGAYRSDDCGTTWVLINDQHQYGNWGOA--IT 711
Db 639 GWSFGFGKASSTGSYVVIYGFPTIDGAAGLKFSEDA GTNMQVVISDASHGFGS-GSANVYN 697
Qy 712 GDHANLRVYIGTNGRGIVYGDIGAPSG 740
Db 698 GDLQTYGRVFRGHERPGHLLRQSOREPAG 726

RESULT 9
US-10-155-400-7
; Sequence 7, Application US/10155400
; Publication No. US20030108988A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; FILE REFERENCE: CELULOLYTICUS
; CURRENT APPLICATION NUMBER: US/10155,400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent in ver. 2.1
; SEQ ID NO 7
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Aspergillus aculeatus
US-10-155-400-7

Query Match 41.6%; Score 1680; DB 14; Length 726;
Best Local Similarity 46.2%; Pred. No. 4.7e-113;
Matches 346; Conservative 113; Mismatches 258; Indels 32; Gaps 17;

Qy 1 ATTQPYTWSNVAI-GGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVG 59
Db 1 AASQAYTKNVVTVGGGGFTFGIVFNPSAKGVAVARTDIGGAYRLN-SDDTWTPLMDWVG 59

Qy 60 ---NNWNGYGVVSIADPINTNKVAAVGMVYNSWDNDGAILRSSDOGATQITPLPF 116
Db 60 NDTWHDW---GIDALATDPDVTDRVYVAVGMVYTNEDPNVGSILRSTDDGDTWTETKLPF 116

Qy 117 KLGGMPCRGMRGLAVDPNNNDILYFGAPSGKGLWRSTDSGATWSOMTNFPDVGTYIAN 176
Db 117 KVGGMPCRGMRGLAVDPNNKNSILYFGARSGHGLWKSTDYGATWSNVTSTWTGTYQD 176

Qy 177 PTDTTGYQSDITQGVVWVAFDKSSSLGQASKITIFGVADPNPNFWSRDGGATWQAVPGA 236
Db 177 SSST--YTSDDPVIAGVTFDTSGSSGSA TPRI FVGVDAGKSVFKSEDA GTATWAVWSGE 234
```

```
Qy 237 PT-GFIPHKGVDPVNVHLYIATSNVTGPGYDGOSSGDVWKFVSVTSGTWTTRISVPSTDTAN 295
Db 235 PQYGLPHKGVSPBEKTLIYSYANGAGPYDGTNGTVHKYNTISGVWTDISP---TSLAS 291
Qy 236 DYFGVSGLTIDROHPNTIMVATQISWMPDITIIFRSTDGGATWTRIWDWTSYSPNRSURYVL 355
Db 292 TTYGGLSVLDQVPGTLMVAALNCWPDDELIFRSTDSGATWSPIEWNGVPSINYYYS 351
Qy 356 DISAEPMLTFGVQNP-PVPSPKLGMDEMAIDPFNSDRMLYGTGATLYATNDLTWDS 414
Db 352 DISAPWLODTTSDQFPV--RVGMVEALAIIDPFDSNHWLYGTGLTVYGGHDLTNWDS 408
Qy 415 GGQIHIAPMVKLETA VNDLISPPSGAPLISALGDLGGFTHADVAVPSTIFTSPVFTT 474
Db 409 KHNVTVKSLAVGIEEMAVLGLITPPGPGALLSAVGDDGCFYHSDLDAAAPNQAHTPTTGT 468
Qy 475 GTSVDYAEALNPSIIVRAGSPSPQPNDRHVAFTSDGKNWFGQSEPGVTTGGTVAASA 534
Db 469 TNGIDYAGNKPNSNIVRSGASD--DYPT--LALSNFGSTWYADYAASTSTGTGAVALSA 523
Qy 535 DGSRFVWAPGDPGQPVVAVGFGNSWAASQGVPAQAIRSDRVPNPKTFYALSNGTFYRST 594
Db 524 DGDVTLMSSTSGALVSKSQG---TLTAVSSLPSCGAVIASDKSDNTVFGGSAGAIYVSK 580
Qy 595 DGGVTFQVAAAGLPSGAGVGMFHAVPGKEGDLMLAASSGLYHSTNGGSSWSAI-TGVSS 653
Db 581 NTATSFTKTVS-LGSSSTTVNAI-RAHPSIAGDVMASTDGKLWHSTDYGSTTQIGSGVTA 638
Qy 654 AVNVFGKSPGSSYPAPVVGTTGGVTGAYRSDDCGTTWVLINDQHQYGNWGOA--IT 711
Db 639 GWSFGFGKASSTGSYVVIYGFPTIDGAAGLKFSEDA GTNMQVVISDASHGFGS-GSANVYN 697
Qy 712 GDHANLRVYIGTNGRGIVYGDIGAPSG 740
Db 698 GDLQTYGRVFRGHERPGHLLRQSOREPAG 726

RESULT 10
US-10-420-191-2
; Sequence 2, Application US/10420191
; Publication No. US20040067569A1
; GENERAL INFORMATION:
; APPLICANT: No. US20040067569A1ozymes Biotech, Inc.
; APPLICANT: Rev. Michael W.
; APPLICANT: Zaretsky, Elizabeth J.
; APPLICANT: Haas, Jeffrey A.
; TITLE OF INVENTION: Polypeptides Having Xyloglucanase Activity And Nucleic Acids
; FILE REFERENCE: Encoding Same
; CURRENT APPLICATION NUMBER: US/10/420,191
; CURRENT FILING DATE: 2003-04-18
; PRIOR APPLICATION NUMBER: US 60/373,987
; PRIOR FILING DATE: 2002-04-19
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 2
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-10-420-191-2

Query Match 39.1%; Score 1579; DB 15; Length 838;
Best Local Similarity 43.0%; Pred. No. 1.2e-105;
Matches 325; Conservative 123; Mismatches 264; Indels 44; Gaps 20;

Qy 6 YTWNSVAI-GGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWV---GW 60
Db 21 PSWKNVKLGCGGGFVPGIIFHPKTKGVAVARTDIGGLYRLN-ADDSWTATVDGIADNAGW 79

Qy 61 NNWNGYGVVSIADPINTNKVAAVGMVYNSWDNDGAILRSSDOGATQITPLPKLGG 120
Db 80 HNW---GIDAVALDPQDDQKVYAAVGMVYNSWDNPSNGAILIRSDRGATWSTNLPPKVG 136
```







```

; LENGTH: 739
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-1011

Query Match      34.8%; Score 1405; DB 14; Length 739;
Best Local Similarity 40.8%; Pred. No. 4.4e-93;
Matches 308; Conservative 104; Mismatches 257; Indels 86; Gaps 21;

Qy 6 YTNENVAIGGGFVDGIVFNEGAGILYVRTDIGMYRWDAANGRWIPLLDWVGNWNGY 65
Db 42 YRWRNAVIGTGFVTVGLFHPSPVRLAYARDIGAYRWDDRGARWTPLDHLGWDDNWL 101
Qy 66 NGVYSIAADPTNTNKNWAAVGMNTNSDPNDGAILRSSDQATWQITPLPKLGNMNGR 125
Db 102 LGVEAMAVDPHPDRLLYLAQRYVAGWAGN-GAVLRSEDGRATWTRDPLVKGLEDGR 160
Qy 126 GMGERLAVDPNNDNILYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTVIANPTDTTGYQS 185
Db 161 GAGERLLVDPKSSSSLGQASKTIFVGVADPNP-----VFSRDRGGATWQAVCAPGPF 240
Qy 186 DIQGVVWVAFDKSSSSLGQASKTIFVGVADPNP-----VFSRDRGGATWQAVCAPGPF 240
Db 208 SGQGVWF-----LVAAKRTVYAGWGDGDTSGTANLYRTAD-GTTGAVGPRPSGT 257
Qy 241 ---IPHKGVFDPVNHVLIATSNTPGPGYDGGVWKFVSVTSGTWTRISPV---PSTDT 293
Db 258 SAKVPLRAAYDTHRELYVTVDGAPGCGQDGSVKLRTATGATGTEWTPVVKPGTTSBG 317
Qy 294 ANDYFGYSGLTIDRQHPNTIMVATQISWMPDTIIFRSTDDGATWTRIDWTSYPNRSLRY 353
Db 318 SADTFAYGVAVDAARRPCTLVSTNNRWADGTVFRSTDDGRTWTSKLD-----AA 368
Qy 354 VLDISAEPLTFGVQPNPVPSPKLGWDEAMADFPNSDRMLYGTGATLYATNDLTKWD 413
Db 369 VEDVSETPFLDWGDD-----KPKFGMWIQALAVDPYDSQHVYGTGATLYGTRDLKRW- 421
Qy 414 SGGQIHIAPMVKGLEETAVNDLISPPSG-APLISALGDLGGFTHADVTAVPST-1FTSPV 471
Db 422 -----APRIRGLUESAVRQLISPPVGEAHLISGLDYGIMYHERLTASPSRGATNPV 474
Qy 472 FTTGTSVDYAEINPSSIIRAGSFDPSSQPNDRHVAFTDGGKNW--FGSEBPGGVTTGGT 529
Db 475 FGSATGLAAARPAVYVTRTGWGDHNG-----AYSHDGGRTWAPFEAQPIAKAAGP 528
Qy 530 VAASADGSRFWA-----PGDGGQPVVYA-----VGFENSWAASQGVPAANAQIRSDRVNPKT 581
Db 529 IATSDGCGTLLWSFVHWDG-----TTAAHRSTDNAGSWSESVSPKGPATVADPADPTR 583
Qy 582 FYA--LSNGTFRSDGGVTFQPVAAGLPSSGAGVGMFHAVPGKEGDLWLAAS-SGLYHS 638
Db 584 FYAVDFDNGTLVASTDSGRSTFARAGGUP-SGDQFQLVAAPGRSGDGLWLSAKWNGLYRS 642
Qy 639 TNGSSWSAITGVSSAVNVGFKGAPGSYPVAVFVVGITGIGVTAAYRSDDCGTTWVLIND 698
Db 643 TDGGDTFARIDSCWASYTLGFKAADGADYPAIQVGSSTEITAVYRSDDAARTWVRIND 702
Qy 699 DOHQVNGWQAITGTHANLRRVYIGTNGRGIYGD 733
Db 703 DAHQWGMGEAVVGDPRIGHRVYLATNGRGIQYGE 737

RESULT 13
US-10-395-241-14
; Sequence 14, Application US/10395241
; Publication No. US20040038367A1
; GENERAL INFORMATION:
; APPLICANT: YAOI, Katsuro
; APPLICANT: MITSUISHI, Yasushi
; TITLE OF INVENTION: NOVEL XYLOGLYCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE
; TITLE OF INVENTION: ENCODING THE SAME, AND METHOD OF PREPARING THE ENZYME
; FILE REFERENCE: Q73756
; CURRENT APPLICATION NUMBER: US/10/395,241

```

```

; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: JP 2002-83433
; FILING DATE: 2002-03-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 14
; LENGTH: 789
; TYPE: PRT
; ORGANISM: Geotrichum sp. M128
US-10-395-241-14

Query Match      27.7%; Score 1117; DB 15; Length 789;
Best Local Similarity 35.0%; Pred. No. 3.8e-72;
Matches 281; Conservative 124; Mismatches 279; Indels 120; Gaps 31;

Qy 6 YTNENVAIGGGFVDGIVFNEGAGILYVRTDIGMYRWDAANGRWIPLLDWVGNWNGY 65
Db 4 YEFKNVAIVGGGYITGIVAHPKTKDLYARTDIGAYRWDAAGTQWIPLNDLFEADQMN 63
Qy 66 NGVYSIAADPTNTNKNWAAVGMNTNSDPNDGAILRSSDQATWQITPLPKLGNMNGP 124
Db 64 MGTESIALDPNPNDRLLYLAQRYVGDW-----AATVSEDRSQSFYIESPPPMGANDMG 119
Qy 125 RMGERLAVDPNNDNILYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTVIANPTDTTGYQ 184
Db 120 RNNGERLAVNPNFNSNEVWMTGT-EGIKWSSDRAKTWTNVTISIPDAF-----TNGIGYT 172
Qy 185 SDIQGVVWVAFDKSSSSLGQASKTIFVGVADPNPVPFMSRDCGATWQAVPGAPTGI--- 241
Db 173 S-----VIEDP-----ERNGTIYASATAPQ--MYVTHDGGVSEVPVAGQSSMLNRT 218
Qy 242 -----PHKGVFDPVNHVLIATSNTPGPGYDGGVWKFVSVTSGTWTRI- 285
Db 219 TGAPDPKXPASTAPQPMKVALTP--NFLYVTVADYPPGPGWTFGEVWRQNRITSGAMDDIT 276
Qy 286 -----SPVPSDTANDYFGYSGLTIDRQHPNTIMVATQISWMPDTI-----IPRSTDDGAT 336
Db 277 PRVGNSSPAPYNNQTFPAGGFCGLSVDATNPRLVIT-LDRDPGPALDSIYLSLDAGAT 335
Qy 337 WTRIDWTS-----YFNRSLRYVLDISAEPLTFGVQV-----NPPVSP---KLG 379
Db 336 WKDVTQLSSPNSLEGNGWGHPTNAARY-KDGTVPVWLDNFNNGPQWGGYGAPHGTPGLTKFG 394
Qy 380 WMDEAMADFPNSDRMLYGTGATLYATNDLTK-----WDSGGQIHIAAP---MVKGLEETA 431
Db 395 WMSAVLIDPPNPEHMYGTGATLYATNDLTSRVEKDW-----APSWYLOIDGLEENA 446
Qy 432 VNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIPTSFTTGTSTVSDYAEINPSSIIVRA 491
Db 447 ILSLRSPKSGAALLSGIDISGMKHDDLTK-PQKMFQAPQFNLDSIDAAGNFPNVVRA 505
Qy 492 GSFDESSQPNDRHVAFTDGGKNW--FGSEBPGGVTT---GGTVAASADGSRFWAP--G 544
Db 506 GSSGHEYDSACARGAYATDGGDAWTFPTCPFGMANASHYQGSTIAVDASGSIQVWSTKLD 565
Qy 545 DRGQPVVAVGFGNSWAASQGVPA-----NAQIRSDRVNPKTFYALSNGTFYRSRTDGGV 598
Db 566 EQASGFWYSHDYGTKTS-----VPAGDLKAQTANVLSDKVQDGTFFIATDGGKFFVSTDDGGK 621
Qy 599 TFPQVAAAGLPSSGAGVGMFHAVPGKEGDLWL-AASSGLYHSTNGSSWSAI-TGVSSAVN 656
Db 622 SYAAGKAGLVT--GTSLMPAVNPVWAGDVVVPVEGGLFHSDFGASFRVGTANATLVS 679
Qy 657 VGFSGS-----APGSSYPVAVFVVGTT---IGVTAAYRSDDCGTTWVLINDDOHQVNGWQA 710
Db 680 VGAPKSKDGKASAPSAFVINGTDKPGSDIGLYRSDDNGSTWTRVNDQEHNTSG-PTMI 738
Qy 711 TGDHANLRVYIGTNGRGIYGD 734
Db 739 EADPKVYGRVLTNGRGIYADL 762

RESULT 14

```



US-10-395-241-12

; Sequence 12, Application US/10395241  
; Publication No. US20040038367A1  
; GENERAL INFORMATION:  
; APPLICANT: YAOI, Katsuro  
; APPLICANT: MITSUISHI, Yasushi  
; TITLE OF INVENTION: NOVEL XYLOGLUCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE  
; TITLE OF INVENTION: ENCODING THE SAME, AND METHOD OF PREPARING THE ENZYME  
; FILE REFERENCE: Q73756  
; CURRENT APPLICATION NUMBER: US/10/395,241  
; CURRENT FILING DATE: 2003-03-25  
; PRIOR APPLICATION NUMBER: JP 2002-83433  
; PRIOR FILING DATE: 2002-03-25  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 12  
; LENGTH: 812  
; TYPE: PRF  
; ORGANISM: Geotrichum sp. M128  
US-10-395-241-12

Query Match 27.7% Score 1117; DB 15; Length 812;  
Best Local Similarity 35.0%; Pred. No. 4e-72;  
Matches 281; Conservative 124; Mismatches 279; Indels 120; Gaps 31;

Qy 6 YTSNVAIGGGGFDGVIFVNEGAPGILYVRDIDGGMYRDAANGRWITPLLDWVGNNWG 65  
Db :  
Qy 27 YEFKNVAIGGGGITGIVAHPKDLLYARTDIGGARWDAGTSKWIPLNDIFAQDMNI 86  
Db :  
Qy 66 NGVVSTAADPINTNKVAAQVMYT--NSWDPNNDGAILLESSDOGTWQTTPKLGGMMPG 124  
Db :  
Qy 87 MGTEIALDPNNPRDLRLAQRYVGDEW----AAFVYSREDRGQSFTLYESFPFGANDMG 142  
Db :  
Qy 125 RGMERLAVDPNNINILFYGAPSGKGILWRSTDGATWSOMTFPDVTGYIANPTDTTGYQ 184  
Db :  
Qy 143 RNNGERLAVNPFSNEVWMGTRT--EGIKSSDRAKTNVNTSIPDAF-----TNGIGYT 195  
Db :  
Qy 185 SDIQGVVWAFAFDKSSSLGQASKTIIFGVADPNPNVPFWSDGGATGAQVPGATGPI--- 241  
Db :  
Qy 196 S-----VIFDP-----ERNGTIYASATAPOG-MVTHTDGGVSWEPVAGQPSSLNRT 241  
Db :  
Qy 242 -----PHKGVPDPNVHVLVIATSTNGPYDGSGSDGVWKFSVTSCTWTIRI - 285  
Db :  
Qy 242 TGAPPDKKPASIAIAPQMKAALTP--NFLVTVIADYPGPWGVTPEVWRQRNTSAMDDIT 299  
Db :  
Qy 286 -----SPVPSSTDANDYFGYSGLTIDRQHPNTIMVATQISWNPDITI---IFRSTDGGAT 336  
Db :  
Qy 300 PRVGNSSPAVYNQTFPAGGFCGLSVDATNPRLVVIIT-LDRDPGALDSIYLSTDAGAT 358  
Db :  
Qy 337 WTRIDWDTS-----YPNRSURYVLDTSAEPLWTFGVQP-----NPPVPSP---KL 379  
Db :  
Qy 359 WKDVTLQSLSPNLSENGNHGHPNAARY-KDGFVPWLDPFNNGPWGGYGCPHGTEPLTKFG 417  
Db :  
Qy 380 WMDESAMAI DPFNSDRMLYGTGATLYATNDLTK---WDSGQIHIA P-----MVKGLEBETA 431  
Db :  
Qy 418 WMSAVLIDFPNPHLMYGTGATITWTDLSRVEKD W-----AFSYLIQIDIEENA 469  
Db :  
Qy 432 VNDLISPPGAPLISALGDLCGFTHADVAVPSTIFTSPVTTTGTSVDYAEALNPSII VRA 491  
Db :  
Qy 470 ILSLRSPKSGAALLSGIGDISGMKHDDLTK-PQMFAPAQFSLNDSIDAAGNFENVVRA 528  
Db :  
Qy 492 GSFPSPSQPNDRHRVAFSTDGKKW--FGSEPGGVTT---CGTVAASADGSFRFWAP--G 544  
Db :  
Qy 529 GSSGHEYDSACARGAYATDGDAWTIPTTCPGPMNASHYQGSTIAVDASGSI VNSTKLD 589  
Db :  
Qy 545 DPQPVVYAVGFGNSWAASQCVA-----NAQIRSDRVNPKTFFVALSNGTFFYRSTDGCV 598  
Db :  
Qy 589 EQASGPWYSHDYGTWS-----VPAGDLKAQTA NVLSDKVDGTIFYATDGGKFYSTDGGK 644  
Db :  
Qy 599 TFQPVAAAGLPSSGAVGVFMHFAVPCKEGBDLWL-AAASSGLYHSTNGGSSWSAI-TGVSSAVN 656  
Db :  
Qy 645 SYAAKAGAGLV T--CTS LMPA VNPVWAGDVWPVPEGGFLFSTDFCASFTRVGTANATLVS 702  
Db :

Qy	657	VGFCKS-----APGSSYPAFVVGTT--IGGVTCAYESDDCGTTWVLINDDQHOYGNWGQAI	711
Db	703	VCAPKSKSDGKKASAPSAVFTWGTDKPGSDIGLYRSDDNGSTWTRVNDQEHYSG-PTMI	761
Qy	711	TGDHANLRVYIGTNGRGIVYGDI	734
Db	762	EADPKYGVYLTNGRGIVYADL	785
RESULT 15			
US-10-395-241-18			
; Sequence 18, Application US/10395241			
; Publication No. US20040038367A1			
; GENERAL INFORMATION:			
; APPLICANT: YAOI, Katsuro			
; APPLICANT: MITSUTSUSHI, Yasushi			
; TITLE OF INVENTION: NOVEL XYLOGLUCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLY			
; TITLE OF INVENTION: ENCODING THE SAME, AND METHOD OF PREPARING THE ENZYME			
; FILE REFERENCE: Q73756			
; CURRENT APPLICATION NUMBER: US/10/395,241			
; CURRENT FILING DATE: 2003-03-25			
; PRIOR APPLICATION NUMBER: JP 2002-83433			
; PRIOR FILING DATE: 2002-03-25			
; NUMBER OF SEQ ID NOS: 19			
; SOFTWARE: PatentIn version 3.2			
; SEQ ID NO 18			
; LENGTH: 826			
; TYPE: PRT			
; ORGANISM: Artificial Sequence			
; FEATURE:			
; OTHER INFORMATION: Xyloglucan Oligosaccharide-Degrading Enzyme with Histid			
US-10-395-241-18			
Query Match 27.6%; Score 1114; DB 15; Length 826;			
Best Local Similarity 35.0%; Pred. No.6.7e-72;			
Matches 281; Conservative 123; Mismatches 280; Indels 120; Gaps			
Qy	6	YTWSNVAITGGGFFVDGIYPNEGAPGILYVTRDIGMYRWDAAANGRWIPLLDWVGNNWGY	65
Db	5	YEFKNVAITGGGGYITGVAHPKTKDLYARTDIGAYRWDAGTSKWIPLNDFIEAQMNI	64
Qy	66	NGVWSIAADPINTKNVAAVAGMYT--NSWDPNDGAILRSSDDQCATWQITPLPFKLGNNMPG	124
Db	65	MGTESIALDPNNPDRLYLAQRYVGDW---AAFVVSEDRGQSFTIYESPPMGANDWG	120
Qy	125	RMGRLAVDPNNDNILYFGAPSGKGLWRSTDSGATWSQMTNPPDVGTYIANFTDTTGYQ	184
Db	121	RNNGERLAVNPPNSNEVWMTGRT--EGIKWSDRAKTNTNVTISPDFAF-----TNGIGYT	173
Qy	185	SDIQGVWVAFDPKSSSLGQASKTIIFVGVADPNNPVFWSRDGGATWQAVPGAPTGTI---	241
Db	174	S-----VIFDP-----ERNGTIIYASATAFG--MYVTHDGGVSEHPVAGQSSWLNR	219
Qy	242	-----PHRGVDFPNVHLVIATNSNTGGPYDGGSGDVWVKFSVTSGTWTRI-	288
Db	220	TGAFDPKKPASTAPQPMKVALTP--NFLVTVYADYPGPGWGTFGKVMRQNRRTSGAMDDIT	277
Qy	286	-----SPVPSTDTANDFYGSLGILIDROHPNTIMVATQISWWPDI---IFPSTGGAT	330
Db	278	PRVGNSSPAPYNNQTFPAGGFCGLSDVATNPRLVVIT--LDRDPGPAIDISILYSTDAGAT	330
Qy	337	WTRIWDWTS-----YPNRSRLVLDISAEPLMTFGVQP-----NPVPSP---KL	375
Db	337	WKDVTQLSGSPSLNKGWGHPTWAARY-KDGTVPVWMLDFNNGPQWGGYGAPHGTFGLTKFG	395
Qy	380	WMDEAMADIPFNSRMLYGTGATLYATNDLTK-----WDSGGQTHIAP-----MYKGLBETA	433
Db	396	WMSAVLIDPPFNPHELMYGTGATIWATDTLRKEVDW-----APSWYLQIDGIBENA	444
Qy	432	VNDLISPPSGAPLISALGDLGFFTHADVTAVSTIETSPVFTTGTSDVVAELNPSILVRA	499
Db	448	ILSLRSPKSGAALLSGIGDISGMKHDLLTK--PQMFAGAPQFSNLDISDAENFPNVVVRA	508







GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model  
Run on: October 5, 2005, 07:09:01 ; Search time 9741.62 Seconds  
(without alignments)  
3680.789 Million cell updates/sec

Title: US-09-917-376-3  
Perfect score: 4036  
Sequence: 1 ATTOPYTWSNVAIGGGFVD.....YIGTNGRIVYDGGAPSG 740

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues  
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09917376/runat\_04102005\_164328\_26897/app\_query.fasta\_1.2524  
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOPCL=0 -LOPEXT=0  
-UNIT8=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS-human40.cdi -LIST=45  
-DOALIGN=200 -NORM=ext -THR SCORE=pct -THR MAX=100 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09917376 @CNC 1 1 15287 @runat\_04102005\_164328\_26897 -NCPU=6 -ICPU=3  
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.\*  
1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sv.\*  
13: gb.un.\*  
14: gb.vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	4036	100.0	2869	6	AX700058 Sequence
2	2429.5	60.2	299800	1	AP005028 Streptomy
3	2371	58.7	296500	1	AL939128 Streptomy
4	2063	51.1	2823	6	AX565635 Sequence

5	2061.5	51.1	3262	1	AF078038S1
6	2009	49.8	12732	1	AE007608
7	1982	49.1	2950	1	AJ585344 Clostridi
8	1753.5	43.4	2579	8	AB1292929 Agaricus
9	1722	42.7	2745	8	AB015511 Aspergill
10	1620	40.1	3359	8	AY040839 Aspergill
11	1579	39.1	2517	8	AY281371 Hypocrea
12	1560.5	38.7	2724	8	AK110506 Oryza sat
13	1405	34.8	299550	1	AP005031 Streptomy
14	1162.5	28.8	14520	1	AE011809 Xanthomon
15	1160.5	28.8	13278	1	AE001712 Thermotog
16	1145	28.4	11910	1	AE012276 Xanthomon
17	1117	27.7	2367	6	AX924431 Sequence
18	1117	27.7	2646	6	AX924429 Sequence
19	1117	27.7	2716	6	AB089343 Geotrichu
20	1114	27.6	2481	6	AX924435 Sequence
21	979	24.3	2755	8	AB116528 Geotrichu
22	460.5	11.4	590	8	AB1534359 Agaricus
23	368.5	9.1	24208	1	AY281357S2 Unculture
24	325	8.1	202301	1	AE017286 Desulfovi
25	271.5	6.7	47537	1	AY328853 Vibrio pa
26	268.5	6.7	301950	1	AP006570 Gloeobact
27	262	6.5	58128	7	AY357582 Burkholde
28	261	6.5	49575	7	AY328852 Vibrio pa
29	238	5.9	20342	1	AE004613 Pseudomon
30	237	5.9	110000	1	AE017282 03 Continuation (14 o
31	226.5	5.6	110000	1	AE017180_13 Continuation (14 o
32	218.5	5.4	7938	1	AE011079 Methanosa
33	214	5.3	5600	1	AF182518 Pseudomon
34	214	5.3	303226	1	AE016774 Pseudomon
35	206	5.1	348866	1	BX640426 Bordetell
36	205.5	5.1	8833	1	AY608695 Burkholde
37	204.5	5.1	110000	1	RMS91985_10 Continuation (11 o
38	204	5.1	110000	1	AP006618_08 Continuation (9 of
39	199.5	4.9	8722	1	AF546905 Mycobacte
40	198.5	4.9	349497	1	BX640440 Bordetell
41	198	4.9	110000	1	CP000011 01 Continuation (2 of
42	196	4.9	110000	1	BX571966_25 Continuation (26 o
43	193.5	4.8	348257	1	BX640425 Bordetell
44	193	4.8	303400	1	AP006578 Gloeobact
45	192.5	4.8	207212	2	AL671876 Mus muscu

ALIGNMENTS

RESULT 1  
AX700058  
LOCUS AX700058 2869 bp DNA linear PAT 03-APR-2003  
DEFINITION Sequence 2 from Patent WO03012090.  
ACCESSION AX700058  
VERSION AX700058.1 GI:29536021  
KEYWORDS  
SOURCE Acidotherrmus cellulolyticus  
ORGANISM Acidotherrmus cellulolyticus  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Frankineae; Acidotherrmaceae; Acidotherrmus.  
REFERENCE 1  
AUTHORS Ding, S.Y., Adney, W.S., Vinzant, T.B. and Himmel, M.E.  
TITLE Thermal tolerant avicelase from Acidotherrmus cellulolyticus  
JOURNAL Patent: WO 03012090-A 2 13-FEB-2003;  
Midwest Research Institute (US)  
FEATURES  
source  
1. 2869  
/organism="Acidotherrmus cellulolyticus"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:28049"

ORIGIN

Alignment Scores:  
Pred. No.: 6.19e-198 Length: 2869  
Score: 4036.00 Matches: 740  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0



Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0
US-09-917-376-3 (1-740) x AX700058 (1-2869)			
QY	1	AlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAsp	20
DB	139	GGCAGCACTCAGCGGTACACCTGGAGCAACGTGGCGATCGGGGGCGGCTTTGTCGAC	198
QY	21	GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGly	40
DB	199	GGGATCGCTTCAATGAAGGTGCACCGGAATTTCTGTACGTGCGAGCAGCATCGGGGG	258
QY	41	MetTyrArgTTPAspAlaalaasnglyArgTTPleProLeuLeuAspTTPValGlyTTP	60
DB	259	ATGTATCGATGGATGGATGCCCAACGGCGGTGGATCCCTTCTTGATTTGGTGGGATGG	318
QY	61	AsnAsnTTPGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys	80
DB	319	AACAATTTGGGGGTACAAACGGGTCTGTCAGCATTTGCGCAGACCCCGATCAATACAAAG	378
QY	81	ValTTPAlaAlaValGlyMetTyrThrAsnSerTTPAspProAsnAspGlyAlaIleLeu	100
DB	379	GTAATGGCGCGCTCGGAATGTACCAACAGCTGGACCCCAACAGCAGCGAGCGAATCTC	438
QY	101	ArgSerSerAspGlnGlyAlaThrTTPGlnIleThrProLeuProPheLysLeuGlyGly	120
DB	439	CGCTCGTCTGTATCAGGGGCGAACGTGGCAATTAACGCCCTCGCCGTTCAAGCTTGGCGG	498
QY	121	AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle	140
DB	499	AACATGCCGGCGGTGGAAATGGCGGAGCGGCTTGGGTGGATCCAAACAATGACAACTT	558
QY	141	LeuTyrPheGlyAlaProSerGlyLysGlyLeuTTPArgSerThrAspSerGlyAlaThr	160
DB	559	CTGTATTTCCGGCCCGCGAGCGCAAGGGCTCTGGAGAAGCACAGATTCGCGCGCGACC	618
QY	161	TTPSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr	180
DB	619	TGGTCCCAAGTACAGAACTTTCCGACGCTAGGACGATGCAATTTGCAATCCCACTGACACG	678
QY	181	ThrGlyTyrGlnSerAspIleGlnGlyValValTTPValAlaPheAspLysSerSerSer	200
DB	679	ACCGGCTATCAGAGCGATATTCAAGCGCTCGTCTGGGTGCTTTTCGACAAAGTCTTCGTC	738
QY	201	SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal	220
DB	739	TGGCTCGGCAAGCGAGTAAGACCAATTTTGGGGCGTGGCGATCCCAATAATCCGGTTC	798
QY	221	PheTTPSerArgAspGlyGlyAlaThrTTPGlnAlaValProGlyAlaProThrGlyPhe	240
DB	799	TTCTGGAGCAGACAGCGCGCGCGACGTGGCAGCGGTGGCGGGTGGCGCGACCGGCTTC	858
QY	241	IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn	260
DB	859	ATCCCGCACAGGGCGTCTTTGACCCGGTCAACACGCTGCTCTAATTTGCCACACAGCAAT	918
QY	261	ThrGlyGlyProTyrAspGlySerSerGlyAspValTTPLysPheSerValThrSerGly	280
DB	919	ACGGGTGGTTCGTATGACGGGAGCTCCGCGCAGCGTCTGGAATTTCTCGGTGACCTCCGGG	978
QY	281	ThrTTPThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr	300
DB	979	ACATGGACGCGAATCAGCCCGGTACTCTTCGACGACACCGGCCAACCGCATATCG	1038
QY	301	SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer	320
DB	1039	ACGGGCTCTACTATCAGCCCGCAGCACCCGAAACAGATAATGGTGGCAACCCAGATATCG	1098
QY	321	TTPTTPProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTTPThrArgIle	340
DB	1099	TGGTGGCGGACACCAATATCTTTTCGAGCACCCGACGGCGGTGGCGGTGGACGGGATC	1158
QY	341	TTPAspTTPThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGlu	360
DB	1159	TGGGATTTGGACGAGTTATCCCAATCGAAGTTGCGATATGCTTGCATTTCCGCGGAG	1218
QY	361	ProTTPLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTTP	380
DB	1219	CCTTGGCTGACCTTCGGGTACAGCCGAATCCTCCCGTACCAGTCCGAGTCCGAGCTCG	1278
QY	381	MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGly	400
DB	1279	ATGGATGAAGCGATGGCAATCGATCCGTTCAACTCTGATCGATGCTCTACGGAACAGC	1338
QY	401	AlaThrLeuTyrAlaThrAsnAspLeuThrLysTTPAspSerGlyGlyGlnIleHisIle	420
DB	1339	GGCAGTTTGTACGCAACAATGATCTCACGAAGTGGACTCCGGCGGCCAGATTCAATC	1398
QY	421	AlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProProSer	440
DB	1399	GGCGCATGGTCAAAAGGATTGGAGGAGACGGCGTAAACGATCTCATACGCCCGCGTCT	1458
QY	441	GlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThr	460
DB	1459	GGCGCCCGCTCATCAGCGCTCTCGAGACCTCGCGGGCTTCAACCCACGCCGACGTTACT	1518
QY	461	AlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyr	480
DB	1519	GGCGTGCATCGACGATCTTCAGGTCAACGGTGTTCACGACCGCACCGCTCGACTAT	1578
QY	481	AlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSerGlnPro	500
DB	1579	GGGGAATTTGAATCCGCTCGATCATCTTCGCGTGGAAATTCGATCCATCGAGCCAAACG	1638
QY	501	AsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTTPPheGlnGlySerGlu	520
DB	1639	AACGACAGCGACGTCGCGTCTTCGACAGCGCGGCAAGAACTGGTTCACAGCAGCAAA	1698
QY	521	ProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheVal	540
DB	1699	CCTGGCGGGGTGACACGGCGGCGACCGTCCGCGCATCGGCCGACGGCTCTCGTTTCTG	1758
QY	541	TTPAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTTP	560
DB	1759	TGGGCTCCCGCGCATCCCGTCCGCTGTGTGTGTACGACGCTCGGATTTGCAACTCTCTG	1818
QY	561	AlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLys	580
DB	1819	GCTGCTTCGCAAGGTGTTCCCGCAATCCCAAGTCCGCTCAGACCGGGTGAATCCAAAG	1878
QY	581	ThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPhe	600
DB	1879	ACTTTCTATGCCCTATCCAAATGGAACCTTCTATCGAGACACGACGGCGGCTGACATTC	1938
QY	601	GlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaVal	620
DB	1939	CAACCGCTCGCGCGCGTCTTCGAGACGCGTGGCGTGGTGTCTATGTTCCACGCGGTG	1998
QY	621	ProGlyLysGlyAspLeuTTPLeuAlaAlaSerSerSerGlyLeuTyrHisSerThrAsn	640
DB	1999	CCTGGAAGAAGAGCGATCTGTGGCTCGCTGATCGAGCGGGCTTTACCACTCAACCAAT	2058
QY	641	GlyGlySerSerTTPSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGly	660
DB	2059	GGCGGACGAGTGGTCTGCAATACCGCGGTATCTCTCCGGGTGAACGTGGGATTTGGT	2118
QY	661	LysSerAlaProGlyLysSerTyrProAlaValPheValGlyThrIleGlyGlyVal	680
DB	2119	AGTCTGGCGCGGTGGTTCATACCCAGCGCTTTTGTGTCGCGCACGATCGGAGCGGT	2178
QY	681	ThrGlyAlaTyrArgSerAspAspCysGlyThrThrTTPValLeuIleAsnAspGln	700
DB	2179	ACGGGGGGGTACCGCTCCGACGACTGTGGACGACCTGGGTACTGATCAATGATGACCG	2238
QY	701	HisGlnTyrGlyAsnTTPGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgVal	720



Db	2239	CACCAATACGGAATGGGACAGCAATACCGGTGACCACGCAATTTACGGCGGTG	2298		CDS	avermectinins" 16..918 /note="SAV1722" /codon_start=1 /transl_table=1 /product="hypothetical protein" /protein_id="BAC69433.1" /db_xref="GI:29605367"
Qy	721	TyrleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyGlyAlaProSerGly	740			
Db	2299	TACATAGCAGCAACGCGGTGAATTGTATACGGGACATTTGGTGGTCCGTCGGA	2358			
RESULT 2						
AP005028						
LOCUS						
Streptomyces avermitilis genomic DNA, complete genome, section						
8/30.						
ACCESSION	AP005028	BA000030				
VERSION	AP005028.1	GI:29605366				
KEYWORDS						
SOURCE						
ORGANISM						
Streptomyces avermitilis MA-4680						
Streptomyces avermitilis MA-4680						
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;						
Streptomycineae; Streptomycetaceae; Streptomyces.						
REFERENCE						
AUTHORS						
Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C.,						
Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T.,						
Kikuchi,H., Shiba,T., Sakaki,Y. and Hattori,M.						
TITLE						
Genome sequence of an industrial microorganism Streptomyces						
avermilis: deducing the ability of producing secondary						
metabolites						
Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)						
21477403						
11572948						
REFERENCE						
AUTHORS						
Ikeda,H., Ishikawa,J., Hanamoto,A., Shinose,M., Kikuchi,H.,						
Shiba,T., Sakaki,Y., Hattori,M. and Omura,S.						
TITLE						
Complete genome sequence and comparative analysis of the industrial						
microorganism Streptomyces avermitilis						
Nat. Biotechnol. 21 (5), 526-531 (2003)						
22608306						
12632562						
REFERENCE						
AUTHORS						
Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C.,						
Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T.,						
Kushida,N., Director-General of Biotechnology Center, Shiba,T.,						
Sakaki,Y. and Hattori,M.						
TITLE						
Direct Submission						
Submitted (29-MAR-2002) Director-General of Biotechnology Center,						
National Institute of Technology and Evaluation, Biotechnology						
Center; 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan						
(E-mail:bio@nite.go.jp, URL:http://www.bio.nite.go.jp/,						
Tel:81-3-3481-1933, Fax:81-3-3481-8424)						
This work was done in collaboration with Haruo Ikeda(*1), Jun						
Ishikawa(*2), Akiharu Hanamoto(*3), Chigusa Takahashi(*3), Mayumi						
Shinose(*3), Hiroshi Horikawa(*4), Hidekazu Nakazawa(*4), Tomomi						
Osonoe(*4), Norihiro Kushida(*4), Hisashi Kikuchi(*4), Tadayoshi						
Shiba(*5), Yoshiyuki Sakaki(*6,*7), Masahira Hattori(*1,*7)						
and Satoshi Omura(*1,*3).						
Final finishing process and all annotation were done by H. Ikeda						
and J. Ishikawa.						
*1 Kitasato Institute for Life Sciences, Kitasato University						
*2 National Institute of Infectious Diseases						
*3 The Kitasato Institute						
*4 National Institute of Technology and Evaluation						
*5 School of Science, Kitasato University						
*6 Institute of Medical Science, University of Tokyo						
*7 RIKEN, Genomic Sciences Center						
Following url is also available.						
http://avermilis.ls.kitasato-u.ac.jp.						
Location/Qualifiers						
1..299800						
/organism="Streptomyces avermitilis MA-4680"						
/mol_type="genomic DNA"						
/strain="MA-4680"						
/db_xref="taxon:227882"						
/note="This strain is also named as strain: ATCC 31267,						
NCIMB 12804 or NRRL 8165.-synonym: Streptomyces						
FEATURES						
source						
1..299800						
/organism="Streptomyces avermitilis MA-4680"						
/mol_type="genomic DNA"						
/strain="MA-4680"						
/db_xref="taxon:227882"						
/note="This strain is also named as strain: ATCC 31267,						
NCIMB 12804 or NRRL 8165.-synonym: Streptomyces						
avermectinins"						
16..918						
/note="SAV1722"						
/codon_start=1						
/transl_table=1						
/product="hypothetical protein"						
/protein_id="BAC69433.1"						
/db_xref="GI:29605367"						
/translation="MAGRALAVPCADLTPMTSAESDGPGRVDRLLRPVVKMLGEE						
LPAVTERGLAHDRRVALIDRTGKVASAKHPLRLSLTLTVTAQADNAGVMTSDG						
KELWSTDQVDEALSEIVCRPVFLTADPHEALDRADPEDVLRDTRPPVSATGR1						
GGSGPEGTFFDPFAPVHVMTTSLARIALAALANGSAQAERYRPNIVLRTPLTRGTEND						
WIDRTLHLGTGLVLRVARTPCAIPTLAHGTVPADTAALRVLAHNRVTPTESMGFQ						
PCAGVYAOVLRPRGIRLGLDPVHVA"						
complement (1232..3406)						
/note="SAV1723"						
/codon_start=1						
/transl_table=1						
/product="putative mycodextranase"						
/protein_id="BAC69434.1"						
/db_xref="GI:29605368"						
/translation="MHSNLRVVRMSAIGAAVALAAGLLVALAPAAHAAAGADLPPT						
SVEASATTGKIGDHTQGIASASGRQAVRLGACQGVETVFRANAVSVNV						
PDQSGALNVYNGTKLAKTLAVTSKYSDVTSWIAKATHHFFDNARLLGQVQG						
DKIAVEASTVQTVVDVADFEQAAATKAGAVSVTDKAGDSQDSTQAFRNAIAA						
AQGVVWTPGQDYRLTSSNGVQNTLQAGSMHVSVHTSRFIDQSSSGGVHLKDFP						
VTGEVTERVSDNDFNVLGSGSSVGMQLQHLKVLGLWLTGNNLNVVNNRILDT						
TADGLNLNGARGVRVNRNPLRNGDDALAMWSLYSPDTNSSPESNTISQPNLANGTA						
IYGGTDLAVNNLISDTNALSGIAISNQKFLDPPSPFLAGTITVSGNTLVAGMNFN						
WQHPMGALRSDYSIAEANVSITGTTDPSYSAFVFGSGGGHGYATKRVTVDGATV						
RNTGTVVQAESQGAAGKFNVOATGVGAAGIYNCYPYAGSGTFTLTDGGGNAGSSTW						
GDCSAWPOPGQGNPDPTNRNLAKRPATATGSDVYTPGKAVDGTASTWESNTNAP						
QOAWTVDLSSQAVRLVLLKPATAWQARTOTLSVGSGTDSGASTVWASQGYRFPD						
ATGNTATVLPSPGNTLRVLYKLVNTANTGTPAAQSFSEVYLS"						
complement (3464..5065)						
/gene="aglA"						
complement (3464..5065)						
/gene="aglA"						
/note="SAV1724"						
/codon_start=1						
/transl_table=1						
/product="putative alpha-glucosidase"						
/protein_id="BAC69435.1"						
/db_xref="GI:29605369"						
/translation="MCQPIPARIQKDWRTAVIYQVYVRSFADGDTGDTGLAGVRAR						
LPYLGLVDVADWFWNFWYLSPMKGGYVDVADYRAVDPAFGTLAEAEKLTAEARELGIR						
TLVDIVPNHSDQHPWFRALAAAGPGSPERALPHFRGGAHAGELPNDWPSFAGST						
EBVWTQDPDGYLHLFTPEQDNLWNAHVAVROHEHDVLRFWFERGVAGVYIDSAALL						
AKDPDLPIVEGRAPHPFVDRDELHDIYRSWRAIADYEGAVFVGEVWLPDTERFARYL						
RDELHTAFNPSFLSCPDARLRTAIDETLAHPVAGPATWVLCNHDVTRTVRYG						
RTDTGDFAAKAFGTPTDLAAGTRRARAALSLALPGSVYLYQGEELGLPEAIPRE						
SIQDPMHFRSHGVDPGRDGRVPLPWADAPHAGFGGEPMLQPQPARWSA'AAADQAE						
RSMLSLYREALRLPAYADGPTLWLPADGVLAFAFARTEGLMCMVNLADTPAGLPAHS						
RLLSGSLDDQGLRPQDTAAWLRI"						
complement (5097..5972)						
/note="SAV1725"						
/codon_start=1						
/transl_table=1						
/product="putative integral membrane protein"						
/protein_id="BAC69436.1"						
/db_xref="GI:29605370"						
/translation="MSTRILVSPILTSRPRGKAAYWTVFGVLLPALAFLPVYVMVY						
TGAMKSPDEAARTPPSLVPQKWHLSGTVDAWDLMLQPLPAHLWNTVQATGAWLLQLVFC						
TAAAYALSRLKPVFGKLVGGLIATLMVPAQALVLPKLTVDADLPILHTSLINDPLGI						
WLPVANAFLNLLKRFDDQIPRDVLEAAEIDGAGRLRLTWSIVLPMSPVLYGVISF						
ALVAVMQDFLWPLWVFSDDTKQIPISVALVQLSQNIQLTVLIAAMVIAISIPWAMFLVF						
QRHIIAGISAGSTKG"						
complement (5969..6925)						
/note="SAV1726"						
/codon_start=1						
/transl_table=1						
/product="putative integral membrane protein"						
/protein_id="BAC69437.1"						







```
QY 366 GlyValGlnProAsnProValProSerProLysLeuGlyTyrMetAspGluAlaMet 385
Db 176829 GGAGGGAACCCGACCGCCGAGCAGACCCGAAACTCGCTGGATGATCCGAGTCCCTG 176888
QY 386 AlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAla 405
Db 176889 GAGATCGACCGCTTGCAGCTCCGCGCGCATGATACGGAACGCGCGCGCATCTACGCG 176948
QY 406 ThrAsnAspLeuThrLysTyrAspSerGlyGlyGlnIleHisIleAlaProMetValLys 425
Db 176949 ACGGACACCTTGCAGAACTGGGACAGCGGAAGCGTTTCCACATCAAGCCATGCGCGG 177008
QY 426 GlyLeuGluThrAlaValAsnAspLeuIleSerProProSer---GlyAlaProLeu 444
Db 177009 GGCCTGGAGGAGACGGCGCTCAACGACCTCGCTCGCTCCCTCGCGCGCGCCAGCTG 177068
QY 445 IleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSer 464
Db 177069 TTCACGGCGCTCGTGATCGCGCGCTTCGGGCACACGGACCTCACACGCGTGCCTG 177128
QY 465 ThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsn 484
Db 177129 CTGATGTACGCTCGCCGAATTCACACGAGCAGCAGCTCGATACGCGCGGAGCCGAC 177188
QY 485 ProSerIleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHis 504
Db 177189 CCGGACCGGCTGTCGGGTGGCAATCTCGAC-----TCGGGTCCG-----CAT 177233
QY 505 ValAlaPheSerThrAspGlyGlyLysAsnTyrPheGlnGlySerGluProGlyVal 524
Db 177234 GTGGGTTCTTCAGCGGACACCGGCCCACTGTGTCGGGGCGGACCTTCGCGGGGTC 177293
QY 525 ThrThrGlyThrValAlaAlaSerAlaAspGlySerArgPheValTyrAlaProGly 544
Db 177294 AGCGGGGTGGGACGTCGCGCGCGTCCGACGCGAGTCGCTTCGTGTGAGCGCGCGG 177353
QY 545 AspProGlyGlnProValTyrAlaValGlyPheGlyAsnSerTyrAlaAlaSerGln 564
Db 177354 GGCACCGGG-----GTGCAGTACACGACCGCGGTTCGGCACCTCGTGTGCGCGTCCGCG 177407
QY 565 GlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAla 584
Db 177408 GGCCTCCGCGCGCGGCGATCGAGTCCGACCGCGGTCCGACCGCGCGGAGACCTTCTACGCG 177467
QY 585 LeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAla 604
Db 177468 TTCAGTCCGCGAGGTTCTACGTCAGTTCGAGCGCGGGCGGACCTTCACGCGCTCCGCG 177527
QY 605 Ala---GlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLys 623
Db 177528 GCACGGGCGCTCCGAGCGCGGACAGC-----GTGCGCTTCAAGCGCGCTGCCCGCGCAG 177581
QY 624 GluGlyAspLeuTyrPheAlaAlaSerSer-----GlyLeuTyrHisSer 638
Db 177582 AAGGGCGACATCTGGCTGGCGGGCGCGAGCGCGCGGTACGGCTGTGGCACTCG 177641
QY 639 ThrAsnGlyLysSerSerTyrPheAlaIleThrGlyValSerSerAlaValAsnValGly 658
Db 177642 ACGGACGGCGCGCGCGCTTCCACAGCTCGCCACCGTCGACCGGCGGACCATCGCG 177701
QY 659 PheGlyLysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGly 678
Db 177702 TTCGCAAGCGCGCGACCGCGCTCGTACAGACGCTCTACACGCGCGGAAGATCGCG 177761
QY 679 GlyValThrGlyAlaTyrArgSerAspCysGlyThrThrTyrValLeuIleAsnAsp 698
Db 177762 GGTGTGCGGGCATCTTCGCGGTCCGACCAAGGGCGCGAGGTGGACCGCGTCAACGAC 177821
QY 699 AspGlnHisGlnTyrGlyAsnTyrGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArg 718
Db 177822 GATGCCACAGTGGGGTTGGACGGCGCGCGCATCAGCGGTGACCCGAGGTCTACGGG 177881
QY 719 ArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyGlyAlaPro 738
```

---

```
Db 177882 CGCGGTATGTCTGCAGCAACGCGCGCGGATCGTCTACGCGACACCGCGGCTCTTCG 177941
QY 739 SerGly 740
Db 177942 GACGCG 177947
RESULT 3
LOCUS SCO939128/2 296500 bp DNA linear BCT 11-FEB-2003
DEFINITION Streptomyces coelicolor A3(2) complete genome; segment 25/29
ACCESSION AL039128 AL039128 AL039128 AL039128 AL039128 AL039128
VERSION AL039128.1 GI:24413907
KEYWORDS Streptomyces coelicolor A3(2)
SOURCE Streptomyces coelicolor A3(2)
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1
AUTHORS Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L.,
Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H.,
Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M.,
Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S.,
Huang, C.H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neill, S.,
Rabinowitch, E., Rajandream, M.A., Rutherford, K., Rutter, S.,
Seeger, K., Saunders, E., Sharp, S., Squares, R., Squares, S.,
Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B.G.,
Parkhill, J. and Hopwood, D.A.
Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2)
JOURNAL Nature 417 (6885), 141-147 (2002)
MEDLINE 21964410
PUBMED 12000953
REFERENCE 2 (bases 1 to 296500)
AUTHORS Bentley, S.D.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces
sequencing team, Sanger Institute, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk
COMMENT On or before Oct 26, 2002 this sequence version replaced
gi:20520750, gi:20520810, gi:20520753, gi:20520666, gi:20520757,
gi:20520758, gi:20520668, gi:20520759, gi:20520814, gi:20520765,
gi:20520767.
FEATURES
source
1..296500
/organism="Streptomyces coelicolor A3(2)"
/mol_type="genomic DNA"
/strain="A3(2)"
/db_xref="taxon:100226"
/complement(144..443)
/gene="SCO6480"
/notes="synonym: SC9C7.16c"
/complement(144..443)
/gene="SCO6480"
/notes="SC9C7.16c, unknown, len: 99 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC9C7.16c"
/protein_id="CAA22728.1"
/db_xref="GI:4154075"
/db_xref="UniProt/TrEMBL:Q9ZBJ4"
/translation="MTYGCIAISMKTGPFREEVSVLLLSADGLRDAGCELYVVLSE
DEDTTWTVEVTRTHEDHSLPAKDAIGRAMPLTGEFTKQEMSVAGGLG"
complement(517..1563)
/gene="SCO6481"
/notes="synonym: SC9C7.17c"
/complement(517..1563)
/gene="SCO6481"
/notes="SC9C7.17c, conserved hypothetical protein, len:
348aa; similar to several hypothetical proteins eg.
SW1Y0BN MYCTU from Mycobacterium tuberculosis (340 aa)
fasta scores; opt: 998, z-score: 1321.8, E(): 0, (47.68
```



gene	identity in 355 aa overlap)."		gene	TR:086875 (EMBL:U22894) from Streptomyces lividans (547 aa) fasta scores; opt: 1000, z-score: 738.2, E(): 0, (40.0% identity in 497 aa overlap) and TR:086807 (EMBL:AL031031) from Streptomyces coelicolor (916 aa) fasta scores; opt: 358, z-score: 316.9, E(): 2.3e-10, (27.2% identity in 779 aa overlap). Also contains a short region of similarity to TR:Q53897 (EMBL:X60316) AbaA regulatory locus from Streptomyces coelicolor (192 aa) fasta scores; opt: 108, z-score: 223.2, E(): 3.8e-05, (38.5% identity in 135 aa overlap)."	
	/codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="CAA22732.1" /db_xref="GI:4154079" /db_xref="GOA:Q9ZB70" /db_xref="UniProt/TREMBL:Q9ZB70" /translation="MVTGHPRTVTATQILRSQGRQDPAMDHGTEDAPPCHRGAGPD TPADPVTGRVPLAVVYDRLGWSRGARRLFAISKEAIGRAPDOLLVPSVGLP EDAPDGAHDDSYRSTAHDDGPGFGLSSLDGRLSYPAAGRARLTAVRADDVVLWM AVPLVGPQPERLLVLAADAGLRDYADGTDSDAGVAVVERIAPAFALHTDFPGADEL ARPLPEILPSMSVGSARIVAQVLEGLVPLEFSDHVRVPTPDGVPRAERARRRE RAARALAAAGPVPDELDLAEADLEAVYARERLEFNEVSGAIGTSLDLSRTIVEVSR VPRFTDVAGTYLREQVAGGFPDGVPTDTTMMHRAVALEHDEPGRDDVVFVGAM PPFAHTPPFCQMTTGEPLVPRITEEMGHAIASQFEKDIRPLITGRSMLLVPLKART VILGFMILLRHPERPVFNDMDRVTAELAAARAGLVLDNARMYTHQENVAETLQDSMLP HIPARLAGCDIATRYLPGTLLGRVGGWDFDAVLPGARTALVVDVGMHGLNSAAMMG QLRTAVQMGADLIDLPQALRLNLDLQAORLGDTHLATCLYAVYDPITAGSLHLNAGHI PPVLVRADGASLIDLTGAPIGVGVPFVSRYTVAPGRDLVMTCDGLVVRGEDI GVGLATLCEAAHPAASMDDACDTIIRALGAGGGRKDDVALLMARLTGTEPDAVAEWR LARDPAEYGRARAARVREQLYDMGLPKLVAPAEMLVSELVTNARISHARPVALRLVRA DTLLCEVDDDDHLELPALKSAGPDETGRLRVVSTLAREWGASRTGAGTKTVWFELPT ARRR"			7711..8433 /gene="SCO6485" /note="synonym: SC9C7.21" 7711..8433 /gene="SCO6485" /note="SC9C7.21, unknown, len: 240 aa" /codon_start=1 /transl_table=11 /product="hypothetical protein SC9C7.21" /protein_id="CAA22733.1" /db_xref="GI:4154080" /db_xref="GOA:Q9ZB19" /db_xref="UniProt/TREMBL:Q9ZB19" /translation="MSVTSRYRDAMEGFWEASGDRGAVFWDABPTLTAGRHIAHFEP HLTDAALPLIDLCGNGTQTRYFAERFPHVVGADLSAALAHARHADPGGLAAYRQLD ADKAEAAQALHAEIGDANVYVRGLHQCEPDDQRPQPLADALAALLGERGAFVPEPSQA AKTVLMGLAQGPAGPPAKLAPVLRHGLTGPVDEAVPEYLRAGLTVLASGELPLVT TEYTDAGTRIELPSAWVAGTA" 8603..9436 /gene="SCO6486" /note="synonym: SC9C7.22" 8603..9436 /gene="SCO6486" /note="SC9C7.22, possible transport associated protein, len: 277aa; similar to SW:DPFA_BACSU dppa, dipeptide transport protein (274 aa) fasta scores; opt: 448, z-score: 604.3, E(): 2.3e-26, (32.3% identity in 269 aa overlap)."	
CDS	/codon_start=1 /transl_table=11		CDS	Alignment Scores: Pred. No.: 2,37e-110 Length: 296500 Score: 2371.00 Matches: 431 Percent Similarity: 72.29% Conservative: 109 Best Local Similarity: 57.70% Mismatches: 183 Query Match: 58.75% Indels: 24 DB: 1 Gaps: 10	
	US-09-917-376-3 (1-740) x SCO939128 (1-296500)				
gene	identity in 355 aa overlap)."		gene	US-09-917-376-3 (1-740) x SCO939128 (1-296500)	
	/codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="CAA22729.1" /db_xref="GI:4154076" /db_xref="UniProt/TREMBL:Q9ZB73" /translation="MRWENLTDSDHARAADTALFGADAVVTRTDTPEFRGITTHEVR ARSILNRVPGASRMPEFVNPYRGCTHACVCFARKTHSYLDLDTGLGPDQSIVVKV NAPEVLRQLASRWQGEHVMGTNVCYQACGRYRLMPCGIIIEALTERANPESLITK GTILRDLQDLRAARVTEVGVISVSGFTDAELWRTVEGPPAPERLEVRVRLAERG IGCVLMAPVYLPFGDEPAQLRAVTRAAGASVTVLHLRPGAREWFMWMLGRH HPHLYRRYLYADGAYAPKWYQRRITRQVHDLAREYIGIGPARAGMPRAVEPESHRP AEAPAPPEVQLSLI" 1800..2243 /gene="SCO6482" /note="synonym: SC9C7.18" 1800..2243 /gene="SCO6482" /note="SC9C7.18, conserved hypothetical protein, len: 348 aa; similar to two hypothetical proteins e.g. TR:O69685 (EMBL:AL022121) Mycobacterium tuberculosis RV03718c or MV025.066c, 147 aa; fasta scores; opt: 475, z-score: 768.3, E(): 0, (52.1% identity in 146 aa overlap)."			US-09-917-376-3 (1-740) x SCO939128 (1-296500)	
CDS	/codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="CAA22730.1" /db_xref="GI:4154077" /db_xref="UniProt/TREMBL:Q9ZB72" /translation="NAQVATETRIVAADAETVFTLDYSGSTRKLLPHEFSYEVR EGGDEGTLVHWLQATSKVRDCLLEVTPTDGLVEKDRNSMVTTRVTPAGEGK SRVVVLTWQAGGAGVGFFEXTFPAPKGLARIYDALLAKLAEEVK" 2394..4841 /gene="SCO6483" /note="synonym: SC9C7.19" 2394..4841 /gene="SCO6483" /note="SC9C7.19, probable efflux protein, len: 815aa; similar to many, several of which are involved in toxin antibiotic efflux eg. TR:O69070 (EMBL:AF061335) oxytetracycline exporter from Streptomyces rimosus (536 aa) fasta scores; opt: 1020, z-score: 879.9, E(): 0, (36.9% identity in 542 aa overlap)."		CDS	US-09-917-376-3 (1-740) x SCO939128 (1-296500)	
	/codon_start=1 /transl_table=11 /product="putative efflux protein" /protein_id="CAA22731.1" /db_xref="GI:4154078" /db_xref="GOA:Q9ZB71" /db_xref="UniProt/TREMBL:Q9ZB71" /translation="MGGITLVQDPVDPVQDFVVAQAAPGPPVPGDGRDAGLSP RRVLVFGVLMALLAALAEQMVATAPKIVGELGLDRKSWAITAYLLTSTVGLPI YKGLDGLGRKGQFQFALLVFWGALAGRAQTMQLIAFRAVQGVGAGGLMIGVQAI IADI VPRPRARYMGLIGASFGLASVAGPLGCGYFTDHLNWRWCYINVPGFLVTMAV VANVLKLP RPVPRKPLDILGALLAASTCTVLTSMGTEYAMGSRVILGLAAGVA ATLLFLVAEFAPELILRLFRDSIFNVTALVGLVGVAFGAASLYLPTFLQWVDA SATEGLMLPMWGAGVWASISQGLISQTHYRIHPILGSALSVVGMWLSRLDADT SRLHYSIWMVAGIGLVMPVLVLAQNSVYRPTDLGTATANNYFRIGGSGVAHF GTLFAGRLTDLALDRIAEAGVGLPDAEAITPOLVHSLPPALRDAYIGAYDAMPRIF LYLVPLVLTGALLFLKKEPLVSHNAPVTPDPAHAPQPOPAQVPPARSYPVAGGI PVCVTQHPDGTVPRAALTLLDVGQOQIGRSGDDGRYALATPGSGAYVLI AAGG HQPAQVSVTPVRLVDVLLGAGRLAGSVNTADGSPVRDXTVLTINVHGVVAITR SGREGYVITELVAGEYTLAASAPAFRAPALVPSVQAAARETRQDVLGAGVLRGTVR ASGGRVADARVTLTLDAGNVVDTLTGPDGCTFRFVDLSGSEYTVIAAGYPPVATVLQ VAGGGRTERDLQLGHED" 4911..7493 /gene="SCO6484" /note="synonym: SC9C7.20" 4911..7493 /gene="SCO6484" /note="SC9C7.20, conserved hypothetical protein, len: 860aa; similar to many hypothetical proteins eg.			US-09-917-376-3 (1-740) x SCO939128 (1-296500)	



```
QY 6 TyrThrTrpSerAenValAlaIleGlyGlyGlyPheValAaspGlyIleValPheAen 25
DB 70439 TACACCTGGAGAAACGCCCGCATCGAGCGCGCGGTTCGTCCCGCATCGTCTTCAAC 70380
QY 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAasp 45
DB 70379 CGGACCGAGAGGACCTGGCTAGCCCGCACCGACATCGAGGCGGCTACCGCTGGCAG 70320
QY 46 AlaAlaAenGlyArgTrpIleProLeuLeuAaspTrpValGlyTrpAenAasnTrpGlyTyr 65
DB 70319 GAGGAGTCCACACCTCGACCGCGCTCTCGACACACGCTCGCTGGAGCAGCTGGGGGCAC 70260
QY 66 AenglyValValSerIleAlaAaaspProIleAenThrAenLysValTrpAlaAlaVal 85
DB 70259 ACGGCGGTGTGTCTCGCTCGACCGCTCGATCCGACCGCGGCTCTACGCGCGGCTC 70200
QY 86 GlyMetTyrThrAenSerTrpAaspProAasnAaspGlyAlaIleLeuArgSerSerAaspGln 105
DB 70199 GGCACGTACACCAACGACTGGGACCCGACCAACCGCGCGGTGCTCGCTCCGCGACCGG 70140
QY 106 GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAasnMetProGlyArg 125
DB 70139 GTGTGAGCTGGGAGAGCGGACCTGCGTTCAGCTGGCGGCAACATGCGCGCGCGC 70080
QY 126 GlyMetGlyGluArgLeuAlaValAaspProAasnAenAasnIleLeuTyrPheGlyAla 145
DB 70079 GGCATGGGTGAGCGCTGCGCTCGACCGCACGACCAACGACGCTCTGTATCCTGGGCGCC 70020
QY 146 ProSerGlyLysGlyLeuTrpArgSerThrAaspSerGlyAlaThrTrpSerGlnMetThr 165
DB 70019 CCCAGCGCCACGGACTGTGGCGGTCACCGACCGCGGCGTCACTGTGTCCGAGTGAGC 69960
QY 166 AenPheProAaspValGlyThrTyrIleAlaAenProThrAaspThrGlyTyrGlnSer 185
DB 69959 GCGTTCGCCGACCCCGGAATACCGCGAGGACCCGACGACGACGCTCGGCTACCGCTCC 69900
QY 186 AspIleGlnGlyValValTrpValAlaPheApsLysSerSer---SerSerLeuGlyGln 204
DB 69899 GACACACGAGGCATCACCTGGGTACCTTCGACGAGTCGACGGCGCGGCGCGGCGCAG 69840
QY 205 AlaSerLysThrIlePheValGlyValAlaAaspProAasnAenProValPheTrpSerArg 224
DB 69839 GCCACCGCGACCCCTACGTCGGGTTCGGGACGAGGACGCGCTCTACCGCTCGACC 69780
QY 225 AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHisLys 244
DB 69779 GACCGGCGCGACCTGGGAGCGGCTCGCGGCGGACCGACCGGTTACTTGGCCCAACAG 69720
QY 245 GlyValPheAaspProValAasnHisValLeuTyrIleAlaThrSerAasnThrGlyGlyPro 264
DB 69719 GGTGTGCTGACCGGAGAACCGGTACCTGTACTCTCGCTACGACGACACCGCGGCGCG 69660
QY 265 TyrAaspGlySerSerGlyAaspValTrpLysPheSerValThrSerGlyThrTrpThrArg 284
DB 69659 TACGACGGCGCAAGCGCGGCTGTACCGGTACCGGACGCGGACCGGACCGGACCTGGACG 69600
QY 285 IleSerProValProSerThrAaspThrAlaAenAaspTyrPheGlyTyrSerGlyLeuThr 304
DB 69599 ATCAGCCCGCGCGGAGCGGACACCC-----TACTCGGCTTCAGCGGCTTGACC 69549
QY 305 IleAaspArgGlnHisProAenThrIleMetValAlaThrGlnIleSerTrpTrpProAasp 324
DB 69548 GTGACCGGACGCGCGGCGGAGCGGTGATGGCGCGCTACAGCTCGTGGTGGCGGAC 69489
QY 325 ThrIleIlePheArgSerThrAaspGlyAlaThrTrpThrArgIleTrpAaspTrpThr 344
DB 69488 ACGCAGATCTTCGCTCGACGACGACGCGGACATGTTCCGACGCGTGGAGCTACACC 69429
QY 345 SerTyrProAasnArgSerLeuArgTyrValLeuAaspIleSerAlaGluProTrpLeuThr 364
DB 69428 TCGTACCGGACCGGAGAACCGGCTACACGATGGACGCTCTCCTCGTCCCTCGGCTCACC 69369
365 PheGlyValGlnProAasnProValProSerProLysLeuGlyTyrTrpMetAaspGluAla 384
69368 TGGGGCGCAACCCCGCGCGCCCGAGCAGACCCCGAAGCTGGGTGGATGACGAGGCT 69309
QY 385 MetAlaIleAaspProPheAasnSerAaspMetLeuTyrGlyThrGlyAlaThrLeuTyr 404
69308 CTGGAGATCGACCCGCTTCGACTCGGACCGGATGTATCGGCACCGGCGCCACCGTCTAC 69249
QY 405 AlaThrAenAaspLeuThrLysTrp---AaspSerGlyGlyGlnIleHisIleAlaProMet 423
69248 GGCACGAGAACCTCAGCAACTGGGACGACGAGGCGGACGCTTCGCCGCTCGAGCGGATG 69189
QY 424 VallysGlyLeuGluGluThrAlaValAenAaspLeuIleSerProProSerGlyAlaPro 443
69188 GTGCGGGCGCTGGAGGACGCGCTCAACGACCTGGCTCCCGCTCCCGCTCGGTGCGCCG 69129
QY 444 LeuIleSerAlaLeuGlyAaspLeuGlyGlyPheThrHisAlaAaspValThrAlaValPro 463
69128 CTGCTCAGCGCCTCGGTGACGTCGCGGCTTCGCGCACACGACGCTGACCGAGGTCCCG 69069
QY 464 SerThrIlePheThrSerProValPheThrGlyThrSerValAaspTyrAlaGluLeu 483
69068 TCGATGATGTACCTCGCCGAACTTCACCTGACGACGAGCTGGACTTCGCGGAGACG 69009
QY 484 AenProSerIleIleValArgAlaGlySerPheAaspSerSerGlnProAasnAaspArg 503
69008 AGCCCGGACGCTCGGTTCGGGCGGGCATCTCGAC-----TCCGGTCCG----- 68964
QY 504 HisValAlaPheSerThrAaspGlyLysAasnTrpPheGlnGlySerGluProGlyGly 523
68963 CACATCGGCTTCTCCAGGACAAACGGCGCAACTGTTGTCGGGCGCACCGACCGCTCCGG 68904
QY 524 ValThrThrGlyGlyThrValAlaAlaSerAlaAaspGlySerArgPheValTrpAlaPro 543
68903 GTGACGGCGCGGACGCTCGCGGCGGCGGACGCGGCTGCTGTGTGGAGCCG 68844
QY 544 GlyAaspProGlyGlnProValValTyrAlaValGlyPheGlyAasnSerTrpAlaAaSer 563
68843 -----GAGGGTCCGCGCTGACGACCGGCTTGGTACGTTCTGGCAGGCGCTCG 68790
QY 564 GlnGlyValProAlaAasnAlaGlnIleArgSerAaspArgValAasnProLysThrPheTyr 583
68789 ACCGCGCTCCCGCGGCGGATCGTCGAGTCGACCGGCTGAACCCCGCGACCTTCTAC 68730
QY 584 AlaLeuSerAasnGlyThrPheTyrArgSerThrAaspGlyGlyValThrPheGlnProVal 603
68729 GCGTTCAGTCCGGGAGGTCTACGTCAGTACGACGCGGCGGCGGCGGCTTCACCGCTCG 68670
QY 604 AlaAla---GlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGly 622
68669 GCGCGACCGGCTGCGCGCGGTGAC-----GGCGTTCGCTTCAAGGCGCTCGCGGCG 68616
QY 623 LysGluGlyAaspLeuTrpLeuAlaAaSerSer-----GlyLeuTyrHis 637
68615 GGGGAGGCGACGCTGCTGGCTGGCGCGGCGGCGGACCGCGCTGACGCGCTTGGCAG 68556
QY 638 SerThrAasnGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAasnVal 657
68555 TCGACGACGCGGCGGACCTTACACGAGCTCCCGGTGTCGACGCGGCGGACACCGCTC 68496
QY 658 GlyPheGlySerSerAlaProGlySerSerTrpProAlaValPheValValGlyThrIle 677
68495 GCGTTCGCAAGCGCGCGGAGCGTGTACAGACCTCTTACAGCGCGGAGATC 68436
QY 678 GlyGlyValThrGlyAlaTyrArgSerAaspCysGlyThrThrTrpValLeuIleAasn 697
68435 GCGCGGCTCGCGGACATCTTCCGCTCCACGACGCGGCGGCGGACCTGACCGCGGCTCAAC 68376
QY 698 AaspAaspGlnHisGlnTyrGlyAasnTrpGlyGlnAlaIleThrGlyAaspHisAlaAasnLeu 717
68375 GACGACGCCACGAGTGGGCTGACGCGGCGGCGGCTCATCATCCGCGGACCGCGGCTCTAC 68316
QY 718 ArgArgValTyrIleGlyThrAenGlyArgGlyIleValTyrGlyAaspIle----- 734
```



Db 68315 GCGCGGTGACGTGGCGGACCAACGCGCGGTGATCTACGCGCACACCTCGACACG 68256  
 QY 735 ---GlyGlyAlaProSerGly 740  
 Db 68255 GCGCGGGGACCGATCCGGGT 68235

RESULT 4  
 LOCUS AX556535 2823 bp DNA linear PAT 29-NOV-2002  
 DEFINITION Sequence 1 from Patent WO20077242.  
 ACCESSION AX556535  
 VERSION AX556535.1 GI:26000969  
 KEYWORDS  
 SOURCE unidentified  
 ORGANISM unidentified  
 unclassified.  
 1  
 Duffner, F. and Sjoeholm, C.  
 Family 74 xyloglucanases  
 TITLE Patent: WO 02077242-A 1 03-OCT-2002;  
 JOURNAL Novozymes A/S (DK)  
 FEATURES  
 Location/Qualifiers  
 1..2823  
 /organism="unidentified"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:32644"  
 /note="Jonesia sp., DSM 14140"  
 1..2823  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="CAD58415.1"  
 /db\_xref="GI:26000970"  
 /translation="MRUKTVVATALTALLATTAGGIAATSTPMVTPAPPSSPSGYS  
 SWSNVEIVGGGFGVIFVQNDPGLVYARTDIGGAYRLNDSTGRWIPLTDHIGWDDWS  
 HSGILSLATDPVDRNKKVLYLAAGTYSMDWDQNGAILRSADKGETWEKTMLEPFRVGGNM  
 PGRGMBERLADVNNKKVLYLAAGTYSMDWDQNGAILRSADKGETWEKTMLEPFRVGGNM  
 YTGQNGVWVTPDTSKAGOTTQTIYGVADKNNVYSTDGGAATWQVPGQPTGF  
 LAQGLDHKQQLIYATSDTGPGYDGSKGVDRWLDLISGGQWTRISPIPSISSNAPG  
 YSLALDKNDPDIIMVVSQVSWPDMVYSTDGKTSWPIWELNGSQPRTKQSNHDI  
 SGAPWLPFGNTAFPEANPKLGMWTSFEIDPHNSDRFFYGTGAGIYGGTLMNWDKG  
 KKVDTIYKQAQIEETAAQDLAAPPNGINGLYSALADIGGFTHKDIQVFNKYIYNKPH  
 DTVTSIDPAESKPATVVRAGKSIGETTWSVGVSTDAGETWKPATGATPSGVKPGSITV  
 SANASSIVAPEGAPRSTNSGWSWTSVGLPNAOVAASDRVNANTLYGFVDGKFEV  
 STNGASPTAFAFTGPTSGNVNFRVPRGROCHLWLAGVSGSTYGMWRSYDGGKNT  
 KVSQGEQDAVFGKATSSGYPVIFTSKADIGVRGIPRSDDEKTKWRIINDNNHNY  
 WTGASITGDPVGERVYIGTGNGRGIIVGDSSTPPPPQDSSGSGSNTGATPTPETDS  
 DHNTGSGTGTGEPGTGSGTAPDSGNTSPGNTSSCKVRYSTDMGSGFTGAATI  
 TMTSPSPINGWTLAFTYPSGOTISSVNSATQTLSGRNVLKNSGNWNTIPAGGSVEFG  
 FNGAYTSKNTPTTSFSLNGVTCALAD"  
 1..69  
 70..2820  
 /product="unnamed"

CDS

sig\_peptide  
 mat\_peptide  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 7,686-97 Length: 2823  
 Score: 2063.00 Matches: 386  
 Percent Similarity: 67.21% Conservative: 112  
 Best Local Similarity: 52.09% Mismatches: 225  
 Query Match: 51.11% Indels: 18  
 DB: 6 Gaps: 9

US-09-917-376-3 (1-740) x AX556535 (1-2823)

QY 6 TyrThrTrpSerAsnValAlaTleGlyGlyGlyGlyPheValAspGlyIleValPheAsn 25  
 Db 130 TATAGCTGTCGAAGTTCAGATTGTGGCGCGCGGTGTTTGTCCCTGGAATTGCTTTAAC 189  
 QY 26 GlyGlyAlaProGlyIleLeuThrValArgThrAspIleGlyGlyMetTyrArgTrpAsp 45  
 Db 190 CAGAAAGACCCAGGCTTGGTGTATGCCCGCACAGATATTGGTGGAGCGTACCGGCTCAAT 249

QY 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr 65  
 Db 250 GACTCCACCGCGGTGGATCCCTCTACGGACCATATTTGGTTGGATGACTGAGGCCAC 309  
 QY 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85  
 Db 310 TCCGGAATTCTCTCTGCGCACTGACCTGTGTGACCAACCGCGGTGACCTTGCAGCA 369  
 QY 86 GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGln 105  
 Db 370 GCGAGTACTCATGACTGGATCCCAAAACGGTGTATTTTGGTTCGCGGACAAA 429  
 QY 106 GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArg 125  
 Db 430 GCGGAGACCTGGGAAAAGACGATGCTGCGTTCCGTGTGGAGGCAACATCGCGGTGCG 489  
 QY 126 GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuThrPheGlyAla 145  
 Db 490 GGGATGGGAGAACGCTCGCTGACCCAAACAATAATAAAAGTCTGTACTTTGGTGGC 549  
 QY 146 ProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 165  
 Db 550 GAATCGGTAATGGAATATGGAAGTCACTGACTACGGGAAAACCTGGGCAAAAGTGACG 609  
 QY 166 AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGly---TyrGln 184  
 Db 610 AGCTTCCCAATGCTGGTAACCTACGTCGCC-----GATGCTTCTGGTGGGTACACC 660  
 QY 185 SerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGln 204  
 Db 661 GGGCAAAACCAAGGTGTGTGTGGGTAACTTTTGACCCCAACATCAGCGAAAGCTGGTCA 720  
 QY 205 AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArg 224  
 Db 721 ACAACAAACCAATCTACGTAGTGTAGCAGCAACAGCAAGCAACACGTTTATCGATCACC 780  
 QY 225 AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHisLys 244  
 Db 781 GATGCGGGGCAACATGCAACCGCTCCTCGCAACCCACCGATTCTCTGGCGGCAAAAG 840  
 QY 245 GlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyPro 264  
 Db 841 GCGAGTGTGACCATATAAAGGTGAGCAACCTCTACATCGCTACCTCAGATACCTGGTGC 900  
 QY 265 TyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArg 284  
 Db 901 TATGACGGTCCGAAAGGGGATGTGGCGCTTAGATATTTCAGATGCGCAGTGGACTCGA 960  
 QY 285 IleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThr 304  
 Db 961 ATCAGCCCAATTCCTCTTACC---TCATCGAATTCAGCGTTCGGGTACTCCGGTTAGCC 1017  
 QY 305 IleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAsp 324  
 Db 1018 ATTGACCGCAAAACCTCGACCATCATGTTGTGTTTCCCAAGTGTCTTGGTGGCTGAC 1077  
 QY 325 ThrIleIlePheArgSerThrAspGlyAlaThrTrpThrArgIleTrpAspTrpThr 344  
 Db 1078 ATGTAGCTATATATCGCTCCACCGCGGGGAAACCGTGTGCTATCTCGGAACTGAAC 1137  
 QY 345 SerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThr 364  
 Db 1138 GGTTCACAGCCCGGTACAAAGCAGTACACACGAGTACTACGGGGCTCCTCGCTTGAC 1197  
 QY 365 PheGlyValGlnProAsnProValProSerProLysLeuGlyTrpMetAspGluAla 384  
 Db 1198 TTTGGCAATACAGCCCAAGAGCTGAAGCTAACCCCAAACTCGGGTGGATGACCCAGTCC 1257  
 QY 385 MetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyr 404  
 Db 1258 TTCGAGATTGACCGCATAACTCCGACCGCTTCTTCTATGGCACCAGCGCGGCGATCTAT 1317  
 QY 405 AlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetVal 424



```
Db 1318 GCGGAACGACCTCAACAACTGGGATAAGGCAAAAGTTGATATCACTGCTCAAGGCC 1377
      ::::::::::::::::::::
Qy 425 LysGlyLeuGluThrAlaValAsnAspLeuLeuSerProSerGlyAlaProLeu 444
      ::::::::::::::::::::
Db 1378 CAAGGATCGAAGAAACCGACACCGCTTGGCGTCCCTCGGAAACATTTGGCCTA 1437
      ::::::::::::::::::::
Qy 445 IleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSer 464
      ::::::::::::::::::::
Db 1438 TACTCAGCGCTGGCAGACATCGGTGGCTTACCCACAAAGATATCTCTCAGGTGCTTAAC 1497
      ::::::::::::::::::::
Qy 465 Thr---IlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeu 483
      ::::::::::::::::::::
Db 1498 AAGTACTACTACAAAGAACCCACATCAGACCGTCACTTCATGACTTCGCGCAAGC 1557
      ::::::::::::::::::::
Qy 484 AsnProSerIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArg 503
      ::::::::::::::::::::
Db 1558 AAACGGCAACTGTCGTGGTGCAGGCAAA-----TCTATTCTGGGGAACACCTCA 1611
      ::::::::::::::::::::
Qy 504 HisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGly 523
      ::::::::::::::::::::
Db 1612 TGGGTGGGTTCCTCAACTGATGCTGGCAAAACGTGCAAAACCTGGCGCCACCATCAGGG 1671
      ::::::::::::::::::::
Qy 524 ValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro 543
      ::::::::::::::::::::
Db 1672 GTTAAAGGCCCGGCTCCATCACTGTCAGTAAATGCTTCATCATGTGTGGGCCCG 1731
      ::::::::::::::::::::
Qy 544 GlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSer 563
      ::::::::::::::::::::
Db 1732 -----GAGGTGGCGCTCCTCGAGTCCACCAACAGTGGCTCCTCGTGGTCAACGGTG 1785
      ::::::::::::::::::::
Qy 564 GlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyr 583
      ::::::::::::::::::::
Db 1786 AGCGGGCTTCCTCACAATGCCAGTCCGATCAGACCGAGTCAACGCTAACTACTCTTTAC 1845
      ::::::::::::::::::::
Qy 584 AlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProVal 603
      ::::::::::::::::::::
Db 1846 GCCTTCGTGGTGAAGGAAAGTTCTATCACTCCACAAATGGTGGGGCTTCCTTCACGGCGTCA 1905
      ::::::::::::::::::::
Qy 604 Ala---AlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGly 622
      ::::::::::::::::::::
Db 1906 GCTTTCACGGGGTTTCCACACAGTGA-----AAGTTCGTTCCTGGTGGCGTGGCGC 1959
      ::::::::::::::::::::
Qy 623 LysGlyGlyAspLeuTrpLeuAla-----AlaSerSerGlyLeuTyrHis 637
      ::::::::::::::::::::
Db 1960 CGTCAAGTCACTCTCTGGCTTCTGCTGGAGTATCGGGAAGCACCTATGGGATGTGGCGG 2019
      ::::::::::::::::::::
Qy 638 SerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnVal 657
      ::::::::::::::::::::
Db 2020 TCCACTGATGGCGGAAGAACTGGACCAAGGTATCTGGAGTTTCAGGAAGCGCATGCCGTT 2079
      ::::::::::::::::::::
Qy 658 GlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValGlyThrIle 677
      ::::::::::::::::::::
Db 2080 GGTTCGGTAAGCAACCAAGTATGCGGATACCCAGTATTTTACATCCGCCAAGATT 2139
      ::::::::::::::::::::
Qy 678 GlyGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsn 697
      ::::::::::::::::::::
Db 2140 GATGAGTTCGAGGAATTTTCGCTGTGACGATGAGGCAAGACCTGGAAACGCTCAAC 2199
      ::::::::::::::::::::
Qy 698 AspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeu 717
      ::::::::::::::::::::
Db 2200 GACAACCAACACAGTGGTATTGGACTGGCGCATCAATCACCAGTGCAGCCAGATGTCGAA 2259
      ::::::::::::::::::::
Qy 718 ArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyGlyAla 737
      ::::::::::::::::::::
Db 2260 GGGCGCGCTATATTGGTACCAACGCTCGCGCATTTATCGTTGGTATTCCTCCACTCCC 2319
      ::::::::::::::::::::
Qy 738 Pro 738
      ::::::::::::::::::::
Db 2320 CCG 2322
      ::::::::::::::::::::
RESULT 5
AF078038S1
```

```
LOCUS AF078038S1 3262 bp DNA linear BCT 11-FEB-2001
DEFINITION Caldicellulosiruptor sp. Tok7B.1 glycosyl hydrolase 5 gene, partial cds.
ACCESSION AF078038
VERSION AF078038.1 GI:12743874
KEYWORDS 1 of 5
SEGMENT
SOURCE
ORGANISM
REFERENCE
AUTHORS Gibbs, M.D., Reeves, R.A., Farrington, G.K., Anderson, P., Williams, D.P. and Bergquist, P.L.
TITLE Multidomain and multifunctional glycosyl hydrolases from the extreme thermophile Caldicellulosiruptor isolate Tok7B.1
JOURNAL Curr. Microbiol. 40 (5), 333-340 (2000)
MEDLINE 20171169
PUBMED 10706665
REFERENCE
AUTHORS Gibbs, M.D., Reeves, R.A., Farrington, G.K., Anderson, P., Williams, D.P. and Bergquist, P.L.
TITLE Direct Submission
JOURNAL Submitted (27-OCT-1999) Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia
FEATURES
source
1..3262
/organism="Caldicellulosiruptor sp. Tok7B.1"
/mol_type="genomic DNA"
/isolate="Tok7B.1"
/db_xref="taxon:80339"
274_..>3262
/note="ORF5; multidomain"
/codon_start=1
/transl_table=1
/product="glycosyl hydrolase 5"
/protein_id="AAK06388.1"
/db_xref="GI:12743879"
/translation="MKRLISLPSILLILNFIABFLQIGAKADTGIAPVYVKNVKI
EGGGFITGIVFNPKENLVYVRDVGAYRSTDDGNTYQLMDVSDNLLGVES
IATDPVDPKXVILACQTYTNSWTNGAILRSTDEGDTPEITPLPXLGGNMPXRL
GSRLAIDPNNRNIIYLGTREGXLMKSEDIYSYDGGKSWFQNTPEQSTSEGTVAAD
HTGVVVVDFDPTSGRPGCKIIYYGVADKTSIYTKDGGQTQWALPQQTGLLPQ
RAKLSDDMLYITYSNTQCPNGDYCEVWRNTKTGSKNI SPMAQDTYFGVGLAV
DAQNPKVVMVAALSSWWPDIYIWRSDGGSTWKCIWENGYPNRTHLYNMDI SAAPWL
NFGXTBPTPEVSPKLGVMVGTLEIDFNSDXMLYGTGAXLYGCDLTLNWDKQNIIT
KYKAIETSTQALISPPVGPLHLSALGDIAGFRHEDLEKAPNWTYVQNMGTDDI
DFAELNPMWRVGNVDKOWNPNTNRIGFSYDGGKSWFQNTPEQSTSEGTVAAD
GSAVWPAKGAICYSTDNGNKWECANVPSEAIYVSDRVPNPKFYAFKNGKFI SAD
KGTFTESPAAGLPISGNFTYVPIEGDIWLNGNMGWHSITDGSYSPVKSIVGVEDAAS
IGFKPAREGETYPAIYIAKINGVRGIFRSDDCDKTWIRINDDKHQFGCANADITGDP
RYGVRVATNGIKWGEIAYNISPSATPTSPSTPTPTPTPTPTPTPTPTPTPTPTPT
PPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT
SAGSIRPWFKIVNGSSSDLSRVKIRYWTVDGKPSQSAVCDMAQIGASNVTFVK
LSSGYSKADYLYLEVGFSSGAGLPQKDTGDIQVRFNKNDWSNINQADDSWLSQMTD
YGENAKVTHLYVDGVLVWGQPGGA"
misc_feature 274..2580
/note="Region: endoglucanase domain"
/evidence=experimental
misc_feature 2791..>3262
/note="type IIIb cellulose binding site"
misc_feature 2581..2790
/note="PT-linker"
ORIGIN
Alignment Scores: 1.08e-96 Length: 3262
Pred. No.: 2061.50 Matches: 387
Score: 65.62% Consistency: 96
Percent Similarity: 52.58% Mismatches: 238
Best Local Similarity: 51.08% Indels: 15
Query Match: 1 Gaps: 9
DB:
```



US-09-917-376-3 (1-740) x AF078038S1 (1-3262)

QY 4 GlnProTyrThrTrpSerAsnValAlaIle---GlyGlyGlyGlyPheValAspGlyIle 22  
 Db 376 CAGCCATATGTTGGAGAAATGTTAAATTTGAAGTGGAGGAGTTTATCAGTGAATA 435  
 QY 23 ValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyr 42  
 Db 436 GTGTTTACCAGGAGAGAAAATCTGTTTACGTTAGGACAGATATAGGAGGCTTAC 495  
 QY 43 ArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsn 62  
 Db 496 CGTAGTACAGACGGTGGCAATACATACAGCAACTTATGGAGCTGGGTAGTTTTCATGAT 555  
 QY 63 TrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysVal--- 81  
 Db 556 TGGAAATTTGCTGGGTGAGAGCAATCGMACAGATCTGTGTGATCTTAATCGWGTATATA 615  
 QY 82 TrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyValAlaIleLeuArg 101  
 Db 616 TTGGCTGCCAGGGGACTTATACCAATAGCTGGACAGATATCAATGGTGTCTATTTTGGGT 675  
 QY 102 SerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsn 121  
 Db 676 TCTACTGACGAAGAGAGATACCTTTTGAATCACTCCACTYCCCTTTTAAATTTGGTGGAAAC 735  
 QY 122 MetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeu 141  
 Db 736 ATGCCRGAAGGAATCTGGGGAAGAGTTAGCTATTGATTCACAAACAATAAGAAATATA 795  
 QY 142 TyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyValAlaThrTrp 161  
 Db 796 TACTTAGAACACAGAGAAGGATGGTGTATGGAAAAGTGAAGTATATGGAGTGTCAATGG 855  
 QY 162 SerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThr 181  
 Db 856 AAAAAGTAAACAAGTTTCTTAACCCAGCCACATATATTGAAGATCCAAACCTGCCCCCAAC 915  
 QY 182 GlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSer 201  
 Db 916 GATTATTTGAACCATATACAGGTGTGCTGGGTGGTATTGATCTTACAAAGTGGTAGG 975  
 QY 202 LeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAlaAspProAsnAsnProValPhe 221  
 Db 976 CCTGTGAGGAAKACAGATATTTATGTGGTGTGTCAGATATTAACCAAGCAATCTAT 1035  
 QY 222 TrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIle 241  
 Db 1036 TATACAAAAGACGGTGGTCAGACATGGCAAGCATTTACCAGCACAGCCAACTGGCTTACTT 1095  
 QY 242 ProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThr 261  
 Db 1096 CCTCAGCGTCAAAATAAGTTCTCT---GATGGGATGTTATATATATAACTTATAGTAACACT 1152  
 QY 262 GlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThr 281  
 Db 1153 CAGGGCCCATATAATGTTGATTATGGAAGATGAGATGAGATCAATACATAAGACCCGAGAA 1212  
 QY 282 TrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSer 301  
 Db 1213 TGGAAAATATAAGTCCCAATGGCTCTCAAGTACT-----TATTTTGGATATGTT 1263  
 QY 302 GlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrp 321  
 Db 1264 GGATTTGGCTGTAGACCCAGCAATCCGAAAGTAGTGTAGTGGTAGCAGCTCTTAGCTCATGG 1323  
 QY 322 TrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrp 341  
 Db 1324 TGGCCAGATACCTATATTATTTGGAAGATATCGATGGTGGCGAAACCGTGAAATATGTTTGG 1383  
 QY 342 AspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluPro 361  
 Db 1384 GAATGGAAATGGCTATCTCTTAATAGAACATTTGATTAATACATGGAATATTTTTCAGCTCT 1443

QY 362 TrpLeuThrPheGly---ValGlnProAsnProValProSerProLysLeuGlyTrp 380  
 Db 1444 TGGCTAACTTTTGGAAWTACTTATCTACACCACTGAAGTAAGCCCAAACTTGGCTGG 1503  
 QY 381 MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGly 400  
 Db 1504 ATGGTTGGAAACACATTTGAATAGACCAATTAATTCAGATARAATGCTTTATCGAACAGCA 1563  
 QY 401 AlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIle 420  
 Db 1564 GCTAAWTGTTATGGATGCGATTTGACAAATTTGGGCAAGGGCAAAATATACCAAT 1623  
 QY 421 AlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProProSer 440  
 Db 1624 AAAGTCAGGCTATAGGAATTTGAAGAACTTCAGTCCAGCACTGATAAGCCACCAGATT 1683  
 QY 441 GlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThr 460  
 Db 1684 GGACCACATTTATTCAGTGCATGGGAGATATAGCGGATTCAGACATGAAGACTTGGAG 1743  
 QY 461 AlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyr 480  
 Db 1744 AAAGCACCTAATTTGGACATATGTTCCAGCTTAATATGGGACTACTACTGATATTGATTT 1803  
 QY 481 AlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSerGlnPro 500  
 Db 1804 GCTGATTTAATCCAAATTTATGTTCTGTAGGGAATGTTGATAAACAGTGGAAATCCA 1863  
 QY 501 AsnAspArgHisAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGly---Ser 519  
 Db 1864 AATCTAACAGAAATTTGTTCTCATATGATGTTGGAATAATCATGTTTCAGGGAATACT 1923  
 QY 520 GluProGlyGlyValThrThrGlyThrValAlaAlaSerAlaAspGlySerArgPhe 539  
 Db 1924 GAACCTCAAGGAACAAGCGAAGTGTACAGTTCGACGTGCTGTCAGATGGTAGTGTCTGT 1983  
 QY 540 ValTrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSer 559  
 Db 1984 GTTTGGGCACCA-----AAAGGCGGAAGTATGCTATTTCTACAGATAACGGAACA 2037  
 QY 560 TrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnPro 579  
 Db 2038 TGGTTGATGCGCAAAATGTAACCTTCAGAACTATTGTTATATTTCTGACAGATGTAATCCA 2097  
 QY 580 LysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThr 599  
 Db 2098 AATAAGTTTATGCAATTTAAGAAATGGCAAAATTTTATATAAGTGTCTCATAAAGGAAACA 2157  
 QY 600 Phe---GlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHis 618  
 Db 2158 TTTATCGAAAGTCTGCTGCTGTTTACCAATATCAGGA-----AATTTCAAA 2205  
 QY 619 AlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSer 638  
 Db 2206 ACTGTTCCAGGTATAGAGGTGATATCTGGCTAGTTGGTAAATAATGGGATGCGCACTCA 2265  
 QY 639 ThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGly 658  
 Db 2266 ACTGATGGTGGATATTCATTTGTAATAATTTTCAAGTGTGAGGATGTCAGCGAGTAGGA 2325  
 QY 659 PheGlyLysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGly 678  
 Db 2326 TTTGGTAAACCTGCTGGAAGGAAACATATCTCGCAATTTATCTTATGCAAAATAAT 2385  
 QY 679 GlyValThrGlyAlaTyrArgSerAspCysGlyThrThrTrpValLeuIleAsnAsp 698  
 Db 2386 GGAGTAAGAGGTATATTTAGATCTGTGATGATAAAGCTGATGAATGAATATGATGAC 2445  
 QY 699 AspGlnHisGlnTyrAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArg 718  
 Db 2446 GACAAACATCAGTTTGGTTGGGAATGCAGACATAACTGGAGATCCAGCGGTGTACGGT 2505











```
RESULT 7
AJ585344
LOCUS
DEFINITION
  381 MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGly 400
  1297 ATGATGGGAGACTTAAAGAAATTCATTAATTAAGTGGATAGAAATGTTTACGGTACAGT 1356
QY
  401 AlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIle 420
  1357 GCAACTCTTATGGAACTGATGATTGACTAATTCGGATAAAGTAAGAAATGACATC 1416
Db
  421 AlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuSerProSer 440
  1417 TCTGTAAAGCAATGGAATTCGAAGATGTCAGTAATGATGCTGTAGTACCAACAAA 1476
QY
  441 GlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThr 460
  1477 GCGCACAACATTTTAAGCCCTGTTGGAGATGACTCGGATTTTATCATGATGACATACT 1536
Db
  461 AlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyr 480
  1537 AAGTTCCAGTAAGTAATGATGACACCACTAATTTAGTCTACACAGCATTTGATTAT 1596
Db
  481 AlaGluLeuAsnProSerIleValArgAlaGlySerPheAspProSerSerGlnPro 500
  1597 GCGAAGAGTTCCTAACTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1656
QY
  501 AsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGln---GlySer 519
  1657 CAAGATAAAGATTGTGGAATTTCTCAGCATGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1716
QY
  520 GluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPhe 539
  1717 AATATATCAGGAGTTTATAAAGCAGGACTGTAGCAGCTGGTCTGCTGCTGCTGCTGCTG 1776
Db
  540 ValTrpAlaProGlyAspProGlyGlnProValValThrAlaValGlyPheGlyAsnSer 559
  1777 GTATGCTCTCCA---GAGGAAGTGTCTAATGCAGCATATAGCACATAATGTAATAAA 1833
QY
  560 TrpAlaAspSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnPro 579
  1834 TGGACACCGTGTCTGCTCTTCCACAGGAGCAAGGTTAGATCAGACCGTGTAAATCCA 1893
Db
  580 LysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThr 599
  1894 AAGAAATCTATGTTCTTAAATGTTAAATCTTATATTAGTACTGATGCGAGTGCACAA 1953
QY
  600 Phe---GlnProValAlaAlaGlyLeuProSerSerGlyValAlaValMetPheHis 618
  1954 TTTACTCAAGTTCACAAACAGTCTCTCTACAAAGGCCAAA---GGAATC---TTTAAG 2007
Db
  619 AlaValProGlyLysGlyAspLeuTrpLeuAla---AlaSerSerGlyLeuTyrHis 637
  2008 ACTGTTATAGTTCATGAAGTGATATTTGGATTGCTGGAGGAAAGACGAGTATGCGAT 2067
QY
  638 SerThrAsnGlyCysSerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnVal 657
  2068 TCAACTGATCTCTGTCACAACTTACAAAGATATCTGTTGTTGTTGTTGTTGTTGTTGTT 2127
Db
  658 GlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValGlyThrIle 677
  2128 GGTCTTGGAAATTCACAAACAGATGATGATATCCAGCTATATATACATGATGATCAAT 2187
QY
  678 GlyGlyValThrGlyAlaTyrArgSerAspCysGlyThrThrTrpValLeuIleAsn 697
  2188 GATGGAACACGCGGAATATTCCTGTTCCAGATGATGAAGTGCACATTTGGGTAAAGATAA 2247
Db
  698 AspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeu 717
  2248 GATGATGCTCACCATAATGTTTCCCGCAGACTACTGCTAATCTGAGATCTTAAACAAATAT 2307
QY
  718 ArgArgValTyrIleGlyThrAsnGlyArgGlyValIleValTyrGlyAspIleGlyGlyAla 737
  2308 GGACGAGTGTCTTGTGGAACATAATGAAGAGGATTTGTTTATGAGACATAGATGTTCT 2367
```

```

AJ585344
Clostridium thermocellum xgha gene for beta-1,4-xyloglucan
hydrolyase, strain F7.
2950 bp DNA linear BCT 11-OCT-2003
ACCESSION
  AJ585344.1 GI:37651952
VERSION
  beta-1,4-xyloglucan hydrolase; xghA gene.
KEYWORDS
  Clostridium thermocellum
SOURCE
  Clostridium thermocellum
ORGANISM
  Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
  Clostridium.
REFERENCE
  1
  Zverlov, V.V.
  New cellulosome components of C. thermocellum
  Unpublished
  2 (bases 1 to 2950)
  Zverlov, V.V.
  Direct Submission
  Submitted (09-OCT-2003) Zverlov V.V., Fachgruppe f. Mikrobielle
  Mikrobiologie, TU Muenchen, Am Hochanger 4, Freising, 85350,
  GERMANY
  Location/Qualifiers
  1..2950
  /organism="Clostridium thermocellum"
  /mol_type="genomic DNA"
  /strain="F7"
  /db_xref="taxon:1515"
  346..2887
  /gene="xghA"
  346..350
  /gene="xghA"
  359..2887
  /gene="xghA"
  /notes="belongs to GH74 and contains a cellulosomal
  dockerin module"
  /codon_start=1
  /evidence=experimental
  /transl_table=11
  /product="beta-1,4-xyloglucan hydrolase"
  /protein_id="CAE51306.1"
  /db_xref="GI:37651953"
  /translation="MVKFTSKIKAAFAAFAAATAIFGPAISQAQVTSVPYKWDNVV
  IGGGFGPFGIVNETEKDLYARADIGGAYRWDPTSTETWIPLLDHFQMDRYSYGVV
  SIATDPVDNRVTIVAGMYNDNLWLPNGAILLRSTDRGETWEKTLIPKMGNGMPGRSM
  GERLAIDPDNDRILYLGTRCGNLWRSTDYGVTSKVSFSPNPGTYIYIDENFDTKDI
  IGVVVVFVKSSSTPGNPTKIYGVADKNESYRSTDGGTWTAKVSPQKGLLPHHG
  VLASGMLYITTGDTGCTGPDGNGKGVQWVKENTRTGWIIDITPTIPVSSDNRFCEAGLA
  VDRONPDILMVTSMNAWPDDEYIFRSTDGATWKNIMWGMYPERILHYEIDISAAPW
  LDWGTSKQLPEINPKLGMIGDIEIDPPNSDRMYVTGATLYGCDNLTDWRGKVKI
  EVKATGIECAVLDLVSPPEGLVSAVGLVGFVHDDLLKVGPKMHPVSVSSGTGID
  YAEVLPNFMAVAKADLYDKKISFYSDGGRNMFQPPNEAPNSVGGGSAVAADAASV
  IWTPEASPAVTTDNGNSKVKCTNLGMAVAVSDRVNGKFPYAFYNGKIFYISTDGLT
  FDTKAPOLPKSVNKIKAVPGKEGHVMAAREGLMSTRTDGGYTFEKLISNVDYTHVVG
  FGKAAQCDYMAIYITGKIDNLVGFPSDDAGKTWVRINDEHGYGAVDTATITGDPVY
  YGRVYIATNGRGIYVGPASDEPVPPTPPQVDVGLVGLNGDNRINSTDLLTLMKRYILK
  SIEDLPVEDDLWAADNGDKINSTDTIYLUKYLQLQAIPELPKK"
ORIGIN
Alignment Scores:
Pred. No.: 1,14e-92 Length: 2950
Score: 1982.00 Matches: 387
Percent Similarity: 65.81% Conservative: 100
Best Local Similarity: 52.30% Mismatches: 229
Query Match: 49.11% Indels: 24
DB: 1 Gaps: 14
US-09-917-376-3 (1-740) x AJ585344 (1-2950)
QY 2 ThrThrGlnProTyrThrTrpSerAsnValAlaIle---GlyGlyGlyPheValAsp 20
Db 458 ACCAGCGTGCCTTACAAATGGGACACAGTGTGTAATCGCGGCGGCGGAGGATTTATCGCG 517
```



QY 21 GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyValArgThrAspIleGly 40  
DB 518 GGTATAGTTTATGAAACGGAAGGATTTGATTTATGACGCTGCGGATATCGGAGA 577  
QY 41 MetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrp 60  
DB 578 GCGTACCGGTGGGATCCTTCGACCGAGACATGGATTCGGTTGCTGGACCATTTCCAAATG 637  
QY 61 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 80  
DB 638 GATGAGTACAGTTATACGGAGTGGAAGATATTCACCGACCCCTGTGGATCCGACCGT 697  
QY 81 ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 100  
DB 698 GTTTACATAGTTGCAGGTATGTATACCAACGATTCGCTTCCTAATATGGAGCAATTCCT 757  
QY 101 ArgSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGly 120  
DB 758 CGCTCAACGGACAGGGAGAACATCGGAAAAACCATACTGCTTTCAGATGGCGGA 817  
QY 121 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 140  
DB 818 AACATGCCGGGAAGATCCATGGGAGAACGCTCTGGCATCGACCCCGAATGACACAGGAT 877  
QY 141 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThr 160  
DB 878 CTTTATCTTGGACACAGTGGGAAACGGACTTTGGAGAGTACCGACTACGGTGTAA 937  
QY 161 TrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 180  
DB 938 TGGTCCAAAGTTGAAAGTTCCCAAAATCCCGAACTTACATTTATGACCCG-----AAT 991  
QY 181 ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSer 200  
DB 992 TTTGATTATACCAAGACATTTATGGAGTAGTCTGGGTGTTTGTGCAAGAGACGAGT 1051  
QY 201 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal 220  
DB 1052 ACACCGGCAACCTCACCAAGACTATATGTTGGTGTGGCTGTGATAAACAAGAGTATT 1111  
QY 221 PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 240  
DB 1112 TACGCGAGTACGACGCGGGGTGTCCACTGGAAAGCAGTTCGCGACAACTTAAGGACTA 1171  
QY 241 IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn 260  
DB 1172 CTTCTCCACCGGGGTTTGTG---GCATCCCAACGGAATGTTGTATATTAACTTATCGTGCAT 1228  
QY 261 ThrGlyGlyProTyrAspGlySer---SerGlyAspValTrpLysPheSerValThrSer 279  
DB 1229 ACCTGCGGTCTTATGACGGCAACGGAAAGGTACAGTTTGGAGTTCAATACACGTACA 1288  
QY 280 GlyThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGly 299  
DB 1289 GGGGAATGGATAGATATCACCCGATACCTTATTCAAGCAGTGCACAT---CGTTCTGCG 1345  
QY 300 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 319  
DB 1346 TTTGACGAGCTTCGAGTCGACGAGGAGAAATCTCGACATTTAATGTTAACTTCCATGAAC 1405  
QY 320 SerTrpTrpProAspThrIlePheArgSerThrAspGlyValAlaThrTrpThrArg 339  
DB 1406 GCGTGGTGGCGGATGAATATATTTTCCGAGTACTGACGGCGGAGCTACATGGGAAGAT 1465  
QY 340 IleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAla 359  
DB 1466 ATCTGGGAATGGGAATGATCTCTGAACGTATACCTGATTCGATTAATGAATAGATATTTCCGCA 1525  
QY 360 GluProTrpLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGly 379  
DB 1526 GCACCGTGGCTGGATTGGGAACTGAGAAAACAGCTGCCGGAATCAATCCGAAACTGGGA 1585  
QY 380 TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr 399

RESULT 8  
AB1292929

DB 1586 TGGATGATAGTGACATAGAGATTGACCGGTTTAATTCGACCGCATGATGATGTATACC 1645  
QY 400 GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHis 419  
DB 1646 GGTGCAACTATCTATGGTTGTGCAATCTTACTGACTGGGACAGAGCGGCAAGTAAAA 1705  
QY 420 IleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAspLeuLysSerProPro 439  
DB 1706 ATCGAGGTAAAGCTACCGGAATAGAGAAATGCGGTATTAGACCTGTTGTAAGCCCGCG 1765  
QY 440 SerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspVal 459  
DB 1766 GAGGGTGCACCGCTTGTAAAGTCAGTTGGCGACCTTGTGCGTTTGTTCATGATGACCTG 1825  
QY 460 ThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAsp 479  
DB 1826 AAAGTTGGTCCGAAAAAATGACGTT---CCTCTTATTCCTTCAGTACGGGAATGAT 1882  
QY 480 TyrAlaGluLeuAsnPro-----SerIleValArgAlaGlySerPheAspPro 496  
DB 1883 TATGCGGAGCTTGTTCGAACTTTATGGCATTTGTCAAAGGCTGATTTGTATGATGA 1942  
QY 497 SerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPhe 516  
DB 1943 -----AAGAAGATTTCTTCTCTTATGACGGAGAGAAATTTGGTTC 1984  
QY 517 Gln---GlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAsp 535  
DB 1985 CAGCACCTAATGAAGACCAAACTCGGTAGGCGCGGTTCGGTTCCGTTGCGAGCCGAT 2044  
QY 536 GlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaValGly 555  
DB 2045 GCATAATCAGTTATTGGACACG---GAAATGCAAGTCCTGCAGTT---ACACGCGAC 2098  
QY 556 PheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAsp 575  
DB 2099 AACGGAAACTCATGGAAGTTTGTACAAATCTTGGTATGGGTGCGGTGGTGCATCCGAC 2158  
QY 576 ArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAsp 595  
DB 2159 CGTGTGACGGTAAATAATTTCTACGCATTTCTATAACGGCAAAATCTATATAAGCACG 2218  
QY 596 GlyGlyValThrPheGlnProVal---AlaAlaGlyLeuProSerSerGlyAlaValGly 614  
DB 2219 GGTGATTTACCTTTACCGATACAAAGGCACCGCAGCTTCCCAAGTCGGTTAAC----- 2272  
QY 615 ValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAla---SerSer 633  
DB 2273 ---AAGATAAAAGCCGTACCGGGCAAGGAAGCACATGTATGGCTTGTCTCAAGAGAAGGC 2329  
QY 634 GlyLeuTyrHisSerThrAsnGlySerSerTrpSerAlaIleThrGlyValSerSer 653  
DB 2330 GGAATTTGGAGGTCCACTGACGCTGGATATACGTTTGAAGAACTCTCCATGTGTACACA 2389  
QY 654 AlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheVal 673  
DB 2390 GCTCATGTGGTAGGCTTCGGAAGGCGACCGGACAGATTCATGCGCATTTACATT 2449  
QY 674 ValGlyThrIleGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrp 693  
DB 2450 ACCGGTAAATTTGACAAATGTTTAGGATCTTCCGTTCCGATGATGTCGCGCAAGACATGG 2509  
QY 694 ValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAsp 713  
DB 2510 GTGCGTATCAACGACGACGACGATATGGCGCTGTTGATATCTGCAATAACAGGTGAC 2569  
QY 714 HisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAsp 733  
DB 2570 CCGAGAGTATACGGAGTGTATATATTCACCAACGGAAGAGGTATTGTTTACGGCGAA 2629











128 CCCGATCGCTTCAATCCCTCCGCAAAAGCGGTGGCGTACGCGCGCAACGGATATCGGA 187  
40 GlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGly 59  
188 GCGCGGTACCGCTGGAAC---TCGAGACACACGTCGACCCCGCTGATGGACTGGGTGGC 244  
60 -----TrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIle 76  
245 AATGATACCTGGCATGACTGG-----GGCATCGATGCGCTTGGCCACGGATCCGGTC 295  
77 AsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAsp 96  
296 GACACGGATCGTGTACGTGCGGTGGCGATGATATACCAATGAATGGACCCCAATGTG 355  
97 GlyAlaIleLeuArgSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPhe 116  
356 GGATCGATCTCGCGTCGACGACGACGAGGTGATACCTGGACTCGAGACGAAGCTGCGGTT 415  
117 LysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsn 136  
416 AAGGTGTGGGAATATCCCGGTCCGGAATGGGTGAGAGATGGCGTCCGCCCGAAC 475  
137 AsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAsp 156  
476 AAGATAGCATTTCTCTATTGGAGCGCGAGTGGCACGGCTTGTGGAGAGACACAGAT 535  
157 SerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTrpIleAlaAsn 176  
536 TATGAGCGACCTGTGCGAATGTCACCTCTTCACGTGGACTGGAACGATATTTCAGGAT 595  
177 ProThrAspThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAsp 196  
596 TCGAGCTCGAC-----TACACTTCGGACCCCTGTGGGTATTGCTTGGGTGACCTTTGAT 649  
197 LysSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspPro 216  
650 TCGACCTCTGGTAGCTCGGGGTCTGCGACCGCTAGGATCTTCTGGAGTTCGGGATGCT 709  
217 AsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAla 236  
710 GGCAGTCTGCTTCAAGTCGGAGGATGCGGGTCTACTTGGCTGGGTGTCTGGAGAG 769  
237 ProThr---GlyPheIleProHisLysGlyValAlaPheAspProValAsnHisValLeuTyr 255  
770 CCTCAGTATGGCTTCTCGCGCACAAAGGGGTCTTCTCCCTGAGGAGAAGACGCTGTAC 829  
256 IleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPhe 275  
830 ATCTCTACGCAATGGCGCTGCTTACGATGGGCACAAACGAAACAGTCCACAAAGTAC 889  
276 SerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAspThrAlaAsn 295  
890 AACATCACACGGAGTGGACGGATATAGTCCC-----ACCTCCCTTGGAGT 940  
296 AspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetVal 315  
941 ACCTACTACGGCTACGGTGTCTGTCGGTTGACCTCCAAAGTCTCGCACCTTGTAGTGTG 1000  
316 AlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAla 335  
1001 GCTGCATCAACTGCTGGTGGCGGATGAGCTCATCTTCGTAGCACTGACTCTGAGCG 1060  
336 ThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeu 355  
1061 ACCTGTGCCCTATTGGAGTGAACGGTTATTCGAGCATCAATTACTACTACTACACTAT 1120  
356 AspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnPro---ProValPro 374  
1121 GATATCAGCAACGCCCTTGGATCCAGGACACGACTTCGACCCGACCAATTCGCCGTT--- 1177  
375 SerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg 394  
1178 -----CGCGTCTGGTGGTGAAGCCCTTGGCATTTGACCATTTGACTCCCAACAC 1231

395 MetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSer 414  
1232 TGGCTCTACGGCACGGCCCTGACAGTATATGGCGGACACGACCTGACCAATTGGGACTCG 1291  
415 GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAsp 434  
1292 AAGCAACATGTACGGTCAAGTCACTCGCGGTGCGTATTGGAGGAGATGGCAGTCTCGGA 1351  
435 LeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPhe 454  
1352 CTGATCACGCCCGCGCGGACCGCGCTGCTCTCGCTTGGCGACGACGGCGGTT 1411  
455 ThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThr 474  
1412 TACCACCTCGGATCTGGACGCGGCCAACCAAGCGTACCACACCGCCTACGGCAGC 1471  
475 GlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValAlaGlySerPhe 494  
1472 ACCAACGGCATCGACTACGCGCGCAACAAAGCCTTCCAACTCGTCCGTCGAGCTAGC 1531  
495 AspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsn 514  
1532 GAC-----GACTATCCGACC-----CTGCTCTGTCGCAACTTCGGAATCGACT 1576  
515 TrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAla 534  
1577 TGTATGACAGACTACGCCCGGTCAACGACACCGCACCGGACGGGTGGTCTTTCGCC 1636  
535 AspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaVal 554  
1637 GATGGTACACGGTCTTCTCATGTCCAGCACCTCGGGCGGTGGTGTGCAAGTCGCAG 1696  
555 GlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSer 574  
1697 GGC-----ACCTTACTGCGGTTCCTCTCTCCCGTCCGCGGGGTGTCATCGCTCC 1747  
575 AspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThr 594  
1748 GACATCCGACAAACACGGTCTTCTATGGTGGATCCGCGGGCGCAATCTAGTCTCGAG 1807  
595 AspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGly 614  
1808 AACACAGGACCTCTCTTCCACCAAGACGTCAGC---CTGGGTCCAGCACCGCCTCAAC 1864  
615 ValMetPheHisAlaValProGlyLysGlyGlyAspLeuTrpLeuAlaAlaSerSerGly 634  
1865 GCCATC---CGCGCGCATCCCTCCATCGCGCGGATGCTCTGGCGGTCCACCGCAAGGCG 1921  
635 LeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIle---ThrGlyValSerSer 653  
1922 CTCTGGCATTCACGGATTACGGTAGTACGTTTCCAGCAGATCCGCGACGGGTCCAGCGG 1981  
654 AlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheVal 673  
1982 GGTGAGCTTGGGTGGCAAGCGCTCTCCACGGGACGCTAGCTCGTATCTACGGG 2041  
674 ValGlyThrIleGlyGlyValThrGlyValArgSerAspAspCysGlyThrThrTrp 693  
2042 TTCTTCCATCCAGCGTGGCGGGGCTGTTCAGAGCGAGGACGCGACGACCACTGG 2101  
694 ValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAla-----IleThr 711  
2102 CAGTGTATCTCGGATGCTCGCACGGATTCGGCTCG---GGGTCCGCAACGCTGTGTAAC 2158  
712 GlyAspHisAlaAsnLeuArgVal-TyrIleGlyThrAsnGlyArgGlyIleValTyr 731  
2159 GGGGACCTGCAGACCTTACGGCGCGGTGTTCTGGGCAACGACGCGCGGGCATTTCTTA 2218  
731 rGlyAspIleGlyGlyAlaProSerGly 740  
2219 CGGCAATCCAGCGGGGACCTGCGCGGC 2246



RESULT 10  
AY040839 3959 bp DNA linear PLN 01-APR-2002  
LOCUS Aspergillus niger endoglucanase C (eglC) gene, complete cds.  
DEFINITION  
ACCESSION AY040839  
VERSION AY040839.1 GI:19879405  
KEYWORDS  
SOURCE  
ORGANISM  
Aspergillus niger  
Aspergillus niger  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
1 (bases 1 to 3959)  
Hasper,A.A., Dekkers,E., van Mil,M., van de Vondervoort,P.J. and de  
Graaff,L.H.  
TITLE  
EglC, a new endoglucanase from Aspergillus niger with major  
activity towards xyloglucan  
Appl. Environ. Microbiol. 68 (4), 1556-1560 (2002)  
JOURNAL  
MEDLINE  
PUBMED 11916668  
REFERENCE  
2 (bases 1 to 3959)  
Hasper,A.A., Dekkers,E. and de Graaff,L.H.  
AUTHORS  
Direct Submission  
TITLE  
Submitted (18-JUN-2001) Molecular Genetics of Industrial  
JOURNAL Micro-organisms, Wageningen University, Dreijenlaan 2, Wageningen  
6703HA, The Netherlands

FEATURES  
source  
1..3959  
/organism="Aspergillus niger"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:5061"  
<1038..>3841  
/gene="eglC"  
join(<1038..1137,1185..1329,1373..1571,1618..1858,  
1908..1961,2007..>3841)  
/gene="eglC"  
/product="endoglucanase C"  
join(1038..1137,1185..1329,1373..1571,1618..1858,  
1908..1961,2007..3841)  
/gene="eglC"  
/codon\_start=1  
/product="endoglucanase C"  
/protein\_id="AAK7727.1"  
/db\_xref="GI:19879406"  
/translation="MRRTFAALVAGLLDSVHAASQAYTWKVVVTGGGGTPTGIVP  
NPSAKGVAYATDGGAYRLNSDDTWTFPLMDWANNNSWHWDGIDAIATDPVDRVYV  
AYGWTNDWPNDSGLRSILKSDQGTWEETKLPFKVGNMGRGVGERLAVDPNDSIL  
YFGARSGLKWSYDGTGNSNVTAFKWTGYSSTYSYSDPVGIAWTFDSTGS  
SGSPTRIFVGVVDTSVFSVSDAGETWTVSGEPYGFPLPHKGLSPSTLYISY  
SNGAGPYDGTNGTVHKYNTSGVWTDISPTSMYDITYGYGLVLDLPVGFVMAALN  
CWMPDELIRWSTDSGTSWSPIMWNGVPSINYYSDISNAPWLDDQVSTDEPPVRVG  
WVVEALIDPDSHLYGTGETIYGGHDLQWDSHNVTIESLAVGIEEMAVGLIT  
PPGPAALLAVDDGGFVHTLUTAPQYIHTPTYSYNTGIDYAGNKPANIVRSGSD  
SDPTIALSSFGSEWADYAAASSTATGQVALSDADATILLNDSGAYRSANSATLSA  
VSSLPAGVIAADKANNTFYCASGSFYLSDTAATFTVTTLGSSSTANAIKRAQS  
LAGDVWSTDTGWLHSTWYKSFPTQISGCTGEGWFGFKPSDGDYVPLFGFFVDG  
VTGLFKETDQGVNQLISDAEHGEGSASNVNGDLQNGYGFVGTNGRGLFYGDPSG  
TLPSATAPASASSTAVKSTSTSTSKVSSSTVSSSTATTITITSSIKSTLITTTKS  
SSSTSTSTSTATGASAYQCGGGSGFTGTPCGSGWCTCTYENEYYSQCKSPIGIATDR  
G"

Alignment Scores:  
Pred. No.: 5,646-74 Length: 3959  
Score: 1620.00 Matches: 352  
Percent Similarity: 55.57% Conservative: 107  
Best Local Similarity: 42.62% Mismatches: 255  
Query Match: 40.14% Indels: 113  
Dbs: 8 Gaps: 20

US-09-917-376-3 (1-740) x AY040839 (1-3959)

QY 1 AlathrThrGlnProTyrThrTrpSerAnValAlaile-----13

Db 1095 GCCCATCCAGGCATATACCTGGAAAGATGTGGTCAC-AGGGGGGTATGACATATCTTGC 1153  
QY 14 -----GlyGlyGlyGlyPheValaspGlyIleVal 23  
Db 1154 AATATGAGGATTGACAAAGCTAAATTGATCAGCGCGGAGGCTTTCAGCGGGGTATCGTC 1213  
QY 24 PheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArg 43  
1214 TTCAACCCCTTCAAGCTAAAGGTGTGCGATATGTCGCACCGACATCGCGCGGTGATATCGG 1273  
QY 44 TrpAspAlaAlaAsnGlyArgTyrPheProLeuLeuAspTyrValGlyTrpAsnAspTyr 63  
1274 TTGAAC---TCGACGACATACATGACCCCGTTCGATGAGTGGGCGCAACAACTCTAACTCG 1330  
QY 64 -----GlyTyr-Asn-----GlyValVa 69  
1331 TACGTAAGCTTCTTACAGATGCTAGGCTCAGTCTCCAGGCACGATTTGGGGCATGTA 1390  
QY 69 LserIleAlaAlaAspProIleAsnThrAsnIysValTrpAlaAlaValGlyMetTyrTh 89  
1391 TGCATCGCGACCGACCGCGTGCATACCGATCGGTATATGTTGCAAGTTGGCATGTACAC 1450  
QY 89 rAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTr 109  
1451 CAACGACTGGGATCCCAATGACGGCTCCATCTCGCGTTCGACGACCGAGCGGATACCTG 1510  
QY 109 pGlnIleThrProLeuProPheLeuGlyAsnMetProGlyArgGlyMetGlyG 129  
1511 GGAAGACAGAAAGCTTCCCTTCAAGTTGGAGGAATATGCCCGTTCGGGTGTGGCGA 1570  
QY 129 u-----ArgLeuAlaVala 134  
1571 GGTAAAGTCTATATCTGGGTTTTGCTAGCATACACTGACCGTCGCGCTCGTCTGG 1630  
QY 134 spProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgS 154  
1631 ATCCGAATGACAAACAGCATCTTGTACTTTGGTCCCGGAGTGGCAATGCTTGTGGAAAA 1690  
QY 154 erThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrI 174  
1691 GCACGAGCTACGCGGAGACATGGTCCATGTGACTGCTTCAAGTGACCGCGACCTTATT 1750  
QY 174 leAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpVala 194  
1751 TCCAAAGTCCAGCTCGAC-----TACACGCTCATCGGTGGGAATTCCTCGGTGCA 1804  
QY 194 laPheAspLysSerSerSerSerLeuGlyGlnAlaSerLysThrIlePheVal----- 211  
1805 CCTTCGACTCGACCTCAGGGTCCAGTGGCTCTCCGACTCCGCGCATCTTTGTCGGTATGT 1864  
QY 212 -----GlyValAlaAspProAsn 217  
1865 GAAAAATTCTAGATATCTCAGGACAGCCCACTGACCATTTACAGGTGTGGGACACTGGA 1924  
QY 218 AsnProValPheTrpSerArgAspGlyGly----- 227  
1925 GAATCGCTTTTGTGTCCGAAGATGCTGGGAAACTTGTAAAGCTTCTCCACGATCGCAAT 1984  
QY 228 -----AlaThrTrpGlnAlaValProGlyAlaProThr---Gly 239  
1985 TAAACCCATCTAACCAACTTAGGACTTGG-----GTATCTGGAGAGCCCATGTACGGC 2038  
QY 240 PheIleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSer 259  
2039 TTCTCCCCCAATAAAGGACATCTCTGTCGCTTCTGAGCACACACTCTACATTTCTTACTCC 2098  
QY 260 AsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSer 279  
2099 AACGGCGCGGTCCATACATGCGCAATGGACCGCTCCACAGTACACATCACCACG 2158  
QY 280 GlyThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGly 299  
2159 GGCGTCTGGACCGACATCAGTCCACCTCCATGACAGACACC-----TACTATGGC 2209



```
QY 300 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 319
DB 2210 TATGCGGAGCTGGCAGTCTCAAGTCCCGGAACTGTGATGGTAGCAGCTCAAC 2269
QY 320 SerTrpTrpProAspThrIlePheArgSerThrAspGlyGlyAlaThrTrpThrArg 339
DB 2270 TGTGTGGTGGCTGATGAGCTGATCTGGCGCAGCAGGATCTTGGGGGACCTGGTCACCC 2329
QY 340 IleTrpAspTrpThrSerTyProAsnArgSerLeuArgTrpValLeuAspIleSerAla 359
DB 2330 ATCTGGGCTTGAAGCGCTATCCAGCATCACTACTACTCTTATGACATCTCCAAC 2389
QY 360 GluProTrpLeuThrPheGlyValGlnProAsnPro---ProValProSerProIleLeu 378
DB 2390 GCCCATCGCTACAGACGACACATCTACAGCAAGAAATTTCCCGTC-----CGCGTC 2440
QY 379 GlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyGly 398
DB 2441 GGTCTGGATGGTGAAGCCCTCGCCATTGATCCCTTCGACTCTGACCATCGCTTTACGGC 2500
QY 399 ThrGlyAlaThrLeuTyAlaThrAsnAspLeuThrIleTrpAspSerGlyGlyGlnIle 418
DB 2501 ACTGCGGAGACCATCTACGGCGGCATGACCTCCAAATTTGGCAGCGAACAACGTA 2560
QY 419 HisIleAlaProMetValIleGlyLeuGluThrAlaValAsnAspLeuIleSerPro 438
DB 2561 ACTATCGAATCTCTGCGTAGGACATCGAAGAAATGGCTGTCTCGGCTCATCACTCC 2620
QY 439 ProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAsp 458
DB 2621 CCAGCGCGGCTGCGCTTCTCTCGGTGCGCAGATGGCGGCTTCTACCACTTCC 2680
QY 459 ValThrAlaValProSerThrIlePheThrSerProValPheThrGlyThrSerVal 478
DB 2681 CTGACACCGCCCTCATCAATATCATCCACCCGACTTACAGTTCAACAAACGGCATC 2740
QY 479 AspTyAlaGluLeuAsnProSerIleValArgAlaGlySerPheAspProSerSer 498
DB 2741 GACTAGCGCGGCAACAGCCCGCTAACATCGTCGCTCTGGATCCAGCGAT----- 2791
QY 499 GlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyIleAsnTrpPheGlnGly 518
DB 2792 -----TCCGACCTACTCTCTCTTATCTCCAGTTTTCGGCGAATCTTGATACCGCGAC 2845
QY 519 SerGluProGlyGlyValThrGlyGlyThrValAlaAlaSerAlaAspGlySerArg 538
DB 2846 TAGCGTGTCTTCTCTACTCGCAGCGCCGCAAGTCGCCCTCTACCGCATCGACACACC 2905
QY 539 PheValTrpAlaProGlyAspProGlyGlnProValValTyAlaValGlyPheGlyAsn 558
DB 2906 ATCTCTCTCATGACAGCGACGGT-----GCCATCGTTCCGCAAACTCGGCT 2953
QY 559 SerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsn 578
DB 2954 ACTCTTAGCGCAGTCTCTCTCTGCGCTCTGGAGCTGTCTATCGCCCTCCGACAAAGTAAAC 3013
QY 579 ProLysThrPheTyAlaLeuSerAsnGlyThrPheTyArgSerThrAspGlyGlyVal 598
DB 3014 AACACCTACTCTACGAGCGCTCCGGCTCTTCTCTTATCTTTCATCCGACACCGCGCG 3073
QY 599 ThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHis 618
DB 3074 ACCTTC---ACTGTACACACACCTCGGAGTAGCAGCAACCGCCAAACCGCATC---CGC 3127
QY 619 AlaValProGlyGlyGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyHisSer 638
DB 3128 GCGCAACCTCTCTCGCAGCGCATGTCTGGTCAGTACAGACACAGAGTCTCTTCCACTCA 3187
QY 639 ThrAsnGlyGlySerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsnVal 657
DB 3188 ACCAATACGCGCAATCTCTTCCACCCAGATTGGAGCGGATGCACAGAGGGGTGGAGTTTC 3247
```

```
QY 658 GlyPheGlyLysSerAlaProGlySerSerTyProAlaValPheValValGlyThrIle 677
DB 3248 GGATTCGGAAGCCCTCTTCGGACGCGACTACCGGCTCTATTTGGATTCTTTACCGTC 3307
QY 678 GlyGlyValThrGlyAlaTyArgSerAspAspCysGlyThrTrpValLeuIleAsn 697
DB 3308 GATGCTGTATCCGGTCTCTTCAAAACCGAAGATCAGGGGTAAATTGGCAGATAAATTC 3367
QY 698 AspAspGlnHisGlnTyGlyAsn---TrpGlyGlnAlaIleThrGlyAspHisAlaAsn 716
DB 3368 GACCGCGAGCACGGTTCGGGTCTCTCGGCAAAATGATGATCAATGGGGATCTTCAGAAC 3427
QY 717 LeuArgArgValTyIleGlyThrAsnGlyArgGlyIleValTyGlyAspIleGlyGly 736
DB 3428 TATGCGAGGTTGTTGGGACAAATGGCGAGGAATCTTCTATGCTGAT----- 3478
QY 737 AlaProSerGly 740
DB 3479 ---CCGAGTGA 3487

RESULT 11
LOCUS AY281371
DEFINITION Hypocrea jecorina strain QM6a Cel74a (cel74a) mRNA, complete cds.
ACCESSION AY281371
VERSION AY281371.1 GI:31747159
KEYWORDS
SOURCE
ORGANISM
Hypocrea jecorina (anamorph: Trichoderma reesei)
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
REFERENCE
1. (bases 1 to 2517)
Foreman, P.K., Brown, D., Dankmeyer, L., Dean, R., Diener, S.,
Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., England, G.J.,
Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchinson, C.,
Olivares, H.A., Teunissen, P.J.M., Yao, J. and Ward, M.
Transcriptional Regulation of Biomass-Degrading Enzymes in the
Filamentous Fungus Trichoderma reesei
J. Biol. Chem. (2003) In press
2. (bases 1 to 2517)
Foreman, P.K., Brown, D., Dankmeyer, L., Dean, R., Diener, S.,
Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., Mitchell, T.,
Olivares, H.A., Teunissen, P.J.M., Yao, J. and Ward, M.
Direct Submission
Submitted (22-APR-2003) Genencor Intl., 925 Page Mill Road, Palo
Alto, CA 94304, USA
FEATURES
Location/Qualifiers
1..2517
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM6a"
/db_xref="taxon:51453"
/notes="anamorph: Trichoderma reesei"
gene
1..2517
/genes="cel74a"
CDS
1..2517
/genes="cel74a"
/notes="putative secreted endoglucanase"
/codon_start=1
/product="Cel74a"
/protein_id="AAP57752.1"
/db_xref="GI:31747160"
/translation="MKVSRVLAIVLGAIVPAHAFAFWKVNKLGCGGFGVPGIIFHPKPT
KGVAARTDIGLYRLNADDSWATVDTGADNAGWNWNGIDAVALDPDDQKVAAVAG
MTYNSDPNGAIIRSDRGATWSFTNLFPKVGNGMPGRGABRLAYVDPANSTIYFG
ARSGNLKSTGCGVTFSKVSFTATGTYIPDSDSNGVNSDKGLMWVTFDSTSTT
CGATSRIFVGTADNITASYVSTNAGTSWAVPGQKYPFHAKALQPKGLVTLVSLNS
DGTGPDYDGLTGSVWRDIAGTWASVPTTYIYSITPKAPWIKNNIDIVTSEPSDG
WMPDAQLPFRSTDSGTTWSPIWASVPTTYIYSITPKAPWIKNNIDIVTSEPSDG
LTKRLGWMLESLEIDPDSNHWLYGTGWTIFGGHDLTNWDRHNVSIOSLADGLEEPS
VODLAPAGGSELLAAVGDNDGFTFASRNDLGTSPQTWATPTWATSTSVYDAGNSVK
SVYRVNCTAGTQOVALSSDGGATWSIDYAADTSMNGGTWAYSDGDTILSTASSGVQ
RSQFQGSFASVSSLPAGVATSKTNSVYFAGSGSTFYVYSKDTGSSFTTRGPKLSAG
```











Db 304 CACAATTGG-----GGAGTGGATGCGCTTCTTGATCCCGAGGACCCCAATGTG 354  
Qy 81 ValTrrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 100  
Db 355 GTGTACTTCGCGCGGGCTGTGTACCAACTCTCTGGGATCTCTGCGCGCGCATCTAC 414  
Qy 101 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLeuLeuGlyGly 120  
Db 415 AAGAGCAGGACAGGGTGGCACTTTCACAGCTCAGCTTCAAGGTCTGGTGGT 474  
Qy 121 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 140  
Db 475 AACATGCTGCGCGCGCATGGCGAGCGCTGCTGCGATCTCGATCCCAAGAACTCCCAAT 534  
Qy 141 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThr 160  
Db 535 GTGTATTTCGCGCGCAGCTCTGGAACCGGCTCTGGAAGAGCACCGGCGCGCAAGACC 594  
Qy 161 TrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 180  
Db 595 TTCAAGAGGTGCACTCTCAAGCTGTGTGGTACCTACATCCCGGTCCCGAGGACG 654  
Qy 181 ThrGlyTyrGlnSerAspIleGlnGlyValTrpValAlaPheAspLysSerSer 200  
Db 655 ACCGGCTACACACGACGACATCAACGGCTGACCTTTGTGACCTTTGTACTCGACCTCCGCA 714  
Qy 201 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnPro 219  
Db 715 ACCATCAATGTGCTGCGCGCATATTCGTAGGCAACCGCGGATACGTCACGCGCTCG 774  
Qy 220 ValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGly 239  
Db 775 CTTTACGTTTCTGAGGATCCCGGACGACATGGAACGGGTCAAGGGCGACCTTGGCAA 834  
Qy 240 PheIleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSer 259  
Db 835 TACTTCTCTCATAGTCAGGCTCGAGCCACCGACCGGCGGCTGTACATCTCTTACGCC 894  
Qy 260 AsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSer 279  
Db 895 GACGAAACGGGCGCTTACGATGTGTCTGCGCTGTCTGGCGATACGCGCTCGCGACC 954  
Qy 280 GlyThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGly 299  
Db 955 GGCACTGGAGCGGACATCACTCTCTCCGGTAGCGGACCTGACC-----TAGGGA 1005  
Qy 300 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 319  
Db 1006 TTCGGCGGCTCTCCGTGGACATGAGAGGACCGAGGAACTCTGGTCTGCGCAGTCTCAAC 1065  
Qy 320 SerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArg 339  
Db 1066 TCTTGGTGGCTGATGCGCAGATCTTCGCGAGCACCGGACTCTGGCAAGACGTTGGTCAAG 1125  
Qy 340 IleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAla 359  
Db 1126 CTCGGCAGTGGGTAACTATCCCGAGATGATCTACTATTACGGCATCTCGACTCCCAAG 1185  
Qy 360 GluProTrpLeuThrPheGlyValGlnProAsnProProValProSerProLysLeu 378  
Db 1186 GCGCCCTGG-----ATCGGCNAGAACTCTCTCGATACCGACTCCAGTGGCTG 1233  
Qy 379 GlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGly 398  
Db 1234 GGCTGGATGATCGAGTCTCTCGAGATCAACCGGACTGACAGCGGACCACTTCTGTACGGC 1293  
Qy 399 ThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIle 418  
Db 1294 ACCGACTGACCCCTCTACGGTGGCGGACGACCTGACCACTGGGACCAACGAGCGCAACATC 1353  
Qy 419 HisIleAlaProMetValLysGlyLeuGluThrAlaValAlaAsnAspLeuIleSerPro 438

Db 1354 ACCATCCAGTCCCTGGCGCAGCGCATCGAGGAGACCGCTGTTCAGGAGGTGGCTTCGGGC 1413  
Qy 439 ProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHis----- 456  
Db 1414 CCGGCGGCGAGCTCATGCTGGTGCACGCGGCGACATCTCCGGCTTCACTTCCGATCG 1473  
Qy 457 ---AlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThrGly 475  
Db 1474 GCGGACAGCTTGGAACTTCAACAAAGCGGTCTGGGACAAACCCCAAGTTCACCACTCA 1533  
Qy 476 ThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAsp 495  
Db 1534 GCCAGCTCGACTACGCGGCAACTTCGCGGGAAGGTGTGCGGTGGGCAAC---GAT 1590  
Qy 496 ProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrp 515  
Db 1591 GCGGCGCTCAG-----ATGTCGCGAGCTCGACCGATGGCGGCTCATATGG 1638  
Qy 516 PheGlnGlySerGluProGlyGlyValThr-----ThrGlyGlyThrValAlaAlaSer 533  
Db 1639 -----AACGTGTACCCCGCGCGCACAAAGACGCGGTGCGGTGTCTACTCT 1692  
Qy 534 AlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTrpAla 553  
Db 1693 GCCGATGTGACAGATCGTGTGTGTCACCGGACCGGTGGC-----GTGTACCGC 1743  
Qy 554 ValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleAsz 573  
Db 1744 TCCGTGAACACGAGGCGAGTTTCTCCGCCGTGTCTGTTCGAGAGCGGAGCTGTCTGGGC 1803  
Qy 574 SerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSer 593  
Db 1804 TCGCAACAGCGCAACAACAGTACTTCTACGCGCATCCGCGGACACTTTTCTACGTCTCC 1863  
Qy 594 ThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaVal 613  
Db 1864 GCTGACGCGCGGTGACCTTT-----GCGCGCGGCTGCTGCTCAGCGCGGCGCACG 1914  
Qy 614 GlyValMet---PheHisAlaValProGlyLysGlyGlyAspLeuTrpLeuAlaAlaSer 632  
Db 1915 TCGATCCGCGACATCGCGCCGACCGAGCGGTGTGGCAAGCTTATGTGTGCGACCAAC 1974  
Qy 633 SerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThr---GlyVal 651  
Db 1975 GCAGCGCTTTCACGAGCAGCGACTTTCGGGACAGCTTTTACCAAGACTTCCGCGGCGCTG 2034  
Qy 652 SerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaVal 671  
Db 2035 AGCGACAGTACAGATCGCGCTTGGCTGGCTCT---GGTTCCAACCTGG---AACCTG 2088  
Qy 672 PheValValGlyThrIleGlyValThrGlyAlaTyrArgSerAspAspCysGlyThr 691  
Db 2089 TACGCTTTGCGACGCGCGCAACCGCGCGAGCTGTACGGCAGCGCGGCGGCGGCTCG 2148  
Qy 692 ThrTrpValLeuIleAsnAspGlnHisGlnTyrGly-----AsnTrpGlyGlnAla 709  
Db 2149 TCCTGGACTGATCTCCAGGCGGACAGCAGGGCTTCGGCATCATCAAC---GGATCGAG 2205  
Qy 710 IleThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIle 729  
Db 2206 CTCGCTGGCAGTGGGAACGTCCCAACAAGGTATACGTCCGGACCAACGCGCGCGGCTC 2265  
Qy 730 ValTyrGlyAsp-----IleGlyGlyAlaProSerSerGly 740  
Db 2266 TTCAGGCGCGAGCGGTTCATCTCCGAGGCGCGCGGCGGACGACGCGCTCTGCT 2316

RESULT 13  
AP005031/c 299550 bp DNA linear BCT 10-MAY-2003  
LOCUS Streptomyces avermitilis genomic DNA, complete genome, section  
DEFINITION 11/30.  
ACCESSION AP005031 BA000030  
VERSION AP005031.1 GI:29606108



## KEYWORDS

SOURCE  
ORGANISM

Streptomyces avermitilis MA-4680  
Streptomyces avermitilis MA-4680  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.

## REFERENCE

AUTHORS

1 Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C.,  
Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T.,  
Kikuchi, H., Shiba, T., Sakaki, Y., and Hattori, M.  
Genome sequence of an industrial microorganism Streptomyces  
avermitilis: deducing the ability of producing secondary  
metabolites

JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)

MEDLINE

21477403

PUBMED

11572948

## REFERENCE

AUTHORS

2 Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H.,  
Shiba, T., Sakaki, Y., Hattori, M., and Omura, S.

Complete genome sequence and comparative analysis of the industrial

TITLE

microorganism Streptomyces avermitilis

JOURNAL

Nat. Biotechnol. 21 (5), 526-531 (2003)

MEDLINE

22608306

PUBMED

12692562

## REFERENCE

AUTHORS

3 Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C.,  
Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T.,  
Kushida, N., Director-General of Biotechnology Center, Shiba, T.,  
Sakaki, Y., and Hattori, M.  
Direct Submission

TITLE

Submitted (29-MAR-2002) Director-General of Biotechnology Center,

JOURNAL

National Institute of Technology and Evaluation, Biotechnology

Center; 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan

(E-mail: bio@nitech.go.jp, URL: http://www.bio.nitech.go.jp/,

Tel: 81-3-3481-1933, Fax: 81-3-3481-8424)

This work was done in collaboration with Haruo Ikeda(\*1), Jun

Ishikawa(\*2), Akiharu Hanamoto(\*3), Chigusa Takahashi(\*3), Mayumi

Shinose(\*3), Hiroshi Horikawa(\*4), Hidekazu Nakazawa(\*4), Tomomi

Osonoe(\*4), Norihiro Kushida(\*4), Hisashi Kikuchi(\*4), Tadayoshi

Shiba(\*5), Yoshiyuki Sakaki(\*6,\*7), Masahira Hattori(\*1,\*7)

and Satoshi Omura(\*1,\*3).

Final finishing process and all annotation were done by H. Ikeda

and J. Ishikawa.

\*1 Kitasato Institute for Life Sciences, Kitasato University

\*2 National Institute of Infectious Diseases

\*3 The Kitasato Institute

\*4 National Institute of Technology and Evaluation

\*5 School of Science, Kitasato University

\*6 Institute of Medical Science, University of Tokyo

\*7 RIKEN, Genomic Sciences Center

Following url is also available.

http://avermitilis.la.kitasato-u.ac.jp.

Location/Qualifiers

## FEATURES

source

1. .299550

/organism="Streptomyces avermitilis MA-4680"

/mol\_type="genomic DNA"

/strain="MA-4680"

/db\_xref="taxon:227882"

/note="This strain is also named as strain: ATCC 31267,

NCIMB 12804 or NRRL 8165.-synonym: Streptomyces

avermitinus"

complement(42. .2936)

/gene="nrd"

complement(42. .2936)

/gene="nrd"

/note="SAV2461"

/codon\_start=1

/transl\_table=11

/product="putative vitamin B12-dependent ribonucleotide

reductase"

/protein\_id="BAC70172.1"

/db\_xref="GI:29606109"

/translation="MTETASGPARSGRAKGTAKGLRIERIHTTTCGVHPYDEVEWARR

DVMTNWRDGSVNFQGVFPDFWSVNAVINVTSKYFRGAVGTQREVSRLIDRI

CDS

/note="SAV2462"

/codon\_start=1

/transl\_table=11

/product="putative regulatory protein"

/protein\_id="BAC70173.1"

/db\_xref="GI:29606110"

/translation="MHCPPCRHPDSRVDSRTTDDGTSIRRRQCDCSRFTTETC

SLMVKRSQVTEPFRKTVINGRKACQGRPVTEDALAQQRVEBAVRATGSARELT

HVGLAILGLPLRELDVAYLRPASVYRFAFSLSDLEFEAAIAELREERQRPDAVDDHE

DAGAEQCTDRGSGGTVEVPATVAD"

4168. .4947

/gene="lexA"

4168. .4947

/gene="lexA"

/note="SAV2463"

/codon\_start=1

/transl\_table=11

/product="putative SOS regulatory protein LexA"

/protein\_id="BAC70174.1"

/db\_xref="GI:29606111"

/translation="MTTADSATITTAQDRSQGRLEPVPHAMNEATSHGPKRSIPGRPP

GRASSGLTDRQRVIEVRQVQRPYPSMREITGOAVGLSSSTSAVHQLMALERK

GFRLDRPHRPAYRVRGSDQSSVQPTDTAGKPAASVYPLVGRRIAGGPILAESVED

VFPLPQLVGDGELFVLKVGDSMIEAAICDGDWVTVRQPVVAENGDIIVAMLDGEAT

VKRFKRGHVWLLPHNSAYQIPGDTEATILGKVVAVLRV"

complement(5110. .7143)

/gene="ding"

complement(5110. .7143)

/gene="ding"

/note="SAV2464"

/codon\_start=1

/transl\_table=11

/product="putative ATP-dependent helicase"

/protein\_id="BAC70175.1"

/db\_xref="GI:29606112"

/translation="MTKPSLPBLLHAAVTAVGTERPGQVTMAEVAEADDSHLLV

QAGTGSIGLYLPAHAGHVRIVATATLALQQLVERDLPTVDALHPLRRERPF

MLKGRSNVLCRLHLEGVPODBEEGLFQFEAAPTSKLQDLRLRDSDETFGD

RDLTPGVSDRAWAQVSVSRRECLGATKAYGAECAEPMRSERAKVAVVYNHALLA

IDAEGAPVLPQHEVLVIDEAHELVSRTVGTVAIGELTPGQVNAVRRAAKLVNEKAD

QQTAAAGFERLMELALPGRLEIPEDLGLTALMALRDAARTVISAIGATDKSVQDED

AVRKAQASVESVDVAERITNGSEWDVVMYERHDFRFGASLVPMSPVSGLLREKLF

DRSVLTATLKLGGDFNGVGSIGLAPGTQGDPLQPMKGVDPVSPFPKQGLYV

AKHLSPARDGDRAVMDLTELIOAGGRITLGLFSSMRAQLAAEELRSRIPFPFIL

LOQETTLGELIKNFAADPTCLFGLTSLMQGVDPGSPCOLVMDKITPFRPDDPLMS

ARQKAVEDAGGNGFMAAATHAALLMAQAGGRIVRASGDRGVAVLQRLATAGVSY

LKASLPDFWYTTDRNQRRSLTAIDEAAKKTTEVTEVAEVAEAAKESGA"

complement(7376. .8509)

/note="SAV2465"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="BAC70176.1"

/db\_xref="GI:29606113"

/translation="MPPTASTDAEPATSDAPRIAVSARSAADSEDTLELRPDELV

ALPAAEAGGEEAAGGAGVREGTAGGERDVGAEPEAGRKDAGPAQGTATA

AFSPPARAGATASADSGSDAVIDPAAFPAAASTVPSASRVPSAPSVTSAPVTAT



CDS	APAPPSLGDTKALAGRAQVSVTRAADLLGRVEDMGPTVTTPAGTFQLAPVPRERDL PLIGRWMDPAVAPRPHDTCTGTHLLIGVADRGGLGSLTLRAVADLVRDRPSCARVAE ADLSLARHYPARPHDTCTGTHLLIGVADRGGLGSLTLRAVADLVRDRPSCARVAE PDLNNTSVSFAFLGAGFRFSAEVDLPDKRAALVWRDRPLRDL		Db	161201 CTCGCGGTGAGCGCATGCCGTCGACCCACCGCCGCGGTCTACCTCGCGGTG 161142		
	/note="(8549..10609) /codon_start=1 /transl_table=1 /product="hypothetical protein" /protein_id="BAC70177.1" /db_xref="GI:29606114"			86 GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGln 105		
	/translation="MNATPTSDGSRSTHERAGHTDQLFTVQGPQAVFASAEQAAPAE TQAPLAPATQVFVEVTPVTVQASSEVPVPGVARESERLRGTATDPLDPEAQTA QAQAVENLLRCWRENGILVAPDDGTLRIPLLATGTALLVPVHWSPTGWHRFGLPYLA DAPQATPADAVTVALLAREAAAGSRAPRAHSAVDHCAVDHSAIDLVRVADSVRRT AVFIDRRARADEPDLFLAABQALLLGHPLPTPKSREGLSEREARLYSPELRGSFA LHWLAVPVLATASWTERGPVPAEQTLARLADGLPMPDGFALPLFPWQIREIR HRPAAALLDAGLRDGLPHGPWHPTSSVTRHTGAPMLKLSLGLRITNSRREN RKELHRGVEVHRLSLGLGEQQAHPQFDIVRPDPAVLAVDDQHGNAVPGLDVMIRHN PFAPTDVSCVAGLSPRPMPQTEQPPVMSRLAEIIRLAGRTGRPGAVATEWF LRYLEOVVRPVLWLDSEAGIALEAHQNTLLLDPEGWPTGGRVDRNOGYVFRYSRA LEEARLPGIHSHTFVSDEVTDERFAYILGINNVGLIGLIGAFGSHLAEQQLLAER RFLADVATGPALRPLTFLRLDSDPVLRCCKANLLTRHLGLDELGVDPDTQSVYVTIAN PLHS"			161141 GGCACCTACGCCACGTCGTGGCGGGCGAAC---GGCGCGGTCTCGCGTCCGAGGACCCG 161085		
	/complement(10775..12229) /note="(SAV2467" /codon_start=1 /transl_table=1 /product="putative class III aminotransferase" /protein_id="BAC70178.1" /db_xref="GI:29606115"			106 GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyAlaAsnMetProGlyArg 125		
	/translation="MAVTKSAVASLKREAGHAHAHEGILRRQSAESAARTYARALP IVPYRGLTTEGADRGVLDCLSGAGTLALGHNHPVVLAIKVLDSGAPLHVLDLA TPVDATFIELRPLAGLADRARVPCGPGATDAVEAAFKLVRATRGSLAFTGA YHMTAGALEASGATDGRVARLPYPQDYPFGIGGREGAELGARWTESLDDVKS PLPAGMLLEPVQEGGVIIPAPDAMLRMRATITADRSIPLIADSVQTVGRTGRFWAV EHSVPDVPVMSKGLPLAVVYRDDLDVWQPGAHAGTFRGNQLAMAAAGTATLA YVRHRLAERAQTLGARMLDOLSLATEFCPCVGEVRGRLMIGVELVRAEDRAVSP GSDPASLDTAGRPAAPALAAVQRECLRGILVELGGRHASVWRLPLPLTISDE QATAVLDKLDAVEARVARGRAERSGRFORSERAG"			161084 GCGGCCACCTGGACCCGCGCACCGCTGACCGTGAAGCTCGCGGCAACGAGGACGCGCGC 161025		
	/complement(10775..12229) /note="(SAV2467" /codon_start=1 /transl_table=1 /product="putative secreted protein" /protein_id="BAC70179.1" /db_xref="GI:29606116"			126 GlyMetGlyIleuArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 145		
	/translation="MRPRLPLFAARRGRSARRRTPSVLGAVGLATATATACDSGD DTKASSESSPASAGDDKIKIPDDIRDKLKERGIDIDKWKNGAWKNWRDMLREAN			161024 GCGCGCGGTGAGCGACTCTCGTCGACCGCGCGACAGCGACACCCCTCTGGCTGGCGACG 160965		
	/complement(10775..12229) /note="(SAV2467" /codon_start=1 /transl_table=1 /product="putative secreted protein" /protein_id="BAC70179.1" /db_xref="GI:29606116"			146 ProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 165		
	/translation="MAVTKSAVASLKREAGHAHAHEGILRRQSAESAARTYARALP IVPYRGLTTEGADRGVLDCLSGAGTLALGHNHPVVLAIKVLDSGAPLHVLDLA TPVDATFIELRPLAGLADRARVPCGPGATDAVEAAFKLVRATRGSLAFTGA YHMTAGALEASGATDGRVARLPYPQDYPFGIGGREGAELGARWTESLDDVKS PLPAGMLLEPVQEGGVIIPAPDAMLRMRATITADRSIPLIADSVQTVGRTGRFWAV EHSVPDVPVMSKGLPLAVVYRDDLDVWQPGAHAGTFRGNQLAMAAAGTATLA YVRHRLAERAQTLGARMLDOLSLATEFCPCVGEVRGRLMIGVELVRAEDRAVSP GSDPASLDTAGRPAAPALAAVQRECLRGILVELGGRHASVWRLPLPLTISDE QATAVLDKLDAVEARVARGRAERSGRFORSERAG"			160964 ---CGGCACGACGGGCTGCTCAAGTCGACCGCGGGCGCCACTTGGGCGACGCGGAC 160908		
	CDS	complement(10775..12229) /note="(SAV2467" /codon_start=1 /transl_table=1 /product="putative class III aminotransferase" /protein_id="BAC70178.1" /db_xref="GI:29606115"		Db	166 AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSer 185	
/translation="MAVTKSAVASLKREAGHAHAHEGILRRQSAESAARTYARALP IVPYRGLTTEGADRGVLDCLSGAGTLALGHNHPVVLAIKVLDSGAPLHVLDLA TPVDATFIELRPLAGLADRARVPCGPGATDAVEAAFKLVRATRGSLAFTGA YHMTAGALEASGATDGRVARLPYPQDYPFGIGGREGAELGARWTESLDDVKS PLPAGMLLEPVQEGGVIIPAPDAMLRMRATITADRSIPLIADSVQTVGRTGRFWAV EHSVPDVPVMSKGLPLAVVYRDDLDVWQPGAHAGTFRGNQLAMAAAGTATLA YVRHRLAERAQTLGARMLDOLSLATEFCPCVGEVRGRLMIGVELVRAEDRAVSP GSDPASLDTAGRPAAPALAAVQRECLRGILVELGGRHASVWRLPLPLTISDE QATAVLDKLDAVEARVARGRAERSGRFORSERAG"			160907 GCCTTCCG-----CGAAGGCGAACTCC 160884			
/complement(10775..12229) /note="(SAV2467" /codon_start=1 /transl_table=1 /product="putative secreted protein" /protein_id="BAC70179.1" /db_xref="GI:29606116"			186 AspileGlnGlyValValTrpValAlaPheAspLysSerSerSerLeuGlyGlnAla 205			
/translation="MAVTKSAVASLKREAGHAHAHEGILRRQSAESAARTYARALP IVPYRGLTTEGADRGVLDCLSGAGTLALGHNHPVVLAIKVLDSGAPLHVLDLA TPVDATFIELRPLAGLADRARVPCGPGATDAVEAAFKLVRATRGSLAFTGA YHMTAGALEASGATDGRVARLPYPQDYPFGIGGREGAELGARWTESLDDVKS PLPAGMLLEPVQEGGVIIPAPDAMLRMRATITADRSIPLIADSVQTVGRTGRFWAV EHSVPDVPVMSKGLPLAVVYRDDLDVWQPGAHAGTFRGNQLAMAAAGTATLA YVRHRLAERAQTLGARMLDOLSLATEFCPCVGEVRGRLMIGVELVRAEDRAVSP GSDPASLDTAGRPAAPALAAVQRECLRGILVELGGRHASVWRLPLPLTISDE QATAVLDKLDAVEARVARGRAERSGRFORSERAG"			160883 TCCGGCGCAGGAGTCTGTTTC-----CTCGTGGCGGCC 160851			
/complement(10775..12229) /note="(SAV2467" /codon_start=1 /transl_table=1 /product="putative secreted protein" /protein_id="BAC70179.1" /db_xref="GI:29606116"			206 SerLysThrIlePheValGlyValAlaAspProAsnAsnPro-----Val 220			
/translation="MAVTKSAVASLKREAGHAHAHEGILRRQSAESAARTYARALP IVPYRGLTTEGADRGVLDCLSGAGTLALGHNHPVVLAIKVLDSGAPLHVLDLA TPVDATFIELRPLAGLADRARVPCGPGATDAVEAAFKLVRATRGSLAFTGA YHMTAGALEASGATDGRVARLPYPQDYPFGIGGREGAELGARWTESLDDVKS PLPAGMLLEPVQEGGVIIPAPDAMLRMRATITADRSIPLIADSVQTVGRTGRFWAV EHSVPDVPVMSKGLPLAVVYRDDLDVWQPGAHAGTFRGNQLAMAAAGTATLA YVRHRLAERAQTLGARMLDOLSLATEFCPCVGEVRGRLMIGVELVRAEDRAVSP GSDPASLDTAGRPAAPALAAVQRECLRGILVELGGRHASVWRLPLPLTISDE QATAVLDKLDAVEARVARGRAERSGRFORSERAG"			160850 GGGCGCACCGTCTACCGCGGTGGGGTGACGGCGCGCACCTCGGCGACGCGGAACCTG 160791			
/complement(10775..12229) /note="(SAV2467" /codon_start=1 /transl_table=1 /product="putative secreted protein" /protein_id="BAC70179.1" /db_xref="GI:29606116"			221 PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 240			
/translation="MAVTKSAVASLKREAGHAHAHEGILRRQSAESAARTYARALP IVPYRGLTTEGADRGVLDCLSGAGTLALGHNHPVVLAIKVLDSGAPLHVLDLA TPVDATFIELRPLAGLADRARVPCGPGATDAVEAAFKLVRATRGSLAFTGA YHMTAGALEASGATDGRVARLPYPQDYPFGIGGREGAELGARWTESLDDVKS PLPAGMLLEPVQEGGVIIPAPDAMLRMRATITADRSIPLIADSVQTVGRTGRFWAV EHSVPDVPVMSKGLPLAVVYRDDLDVWQPGAHAGTFRGNQLAMAAAGTATLA YVRHRLAERAQTLGARMLDOLSLATEFCPCVGEVRGRLMIGVELVRAEDRAVSP GSDPASLDTAGRPAAPALAAVQRECLRGILVELGGRHASVWRLPLPLTISDE QATAVLDKLDAVEARVARGRAERSGRFORSERAG"			160790 TACCACGCGCCGAC---GGCACGACCTGGGGGGCGTCCCGCGCGGCTCCCGGAC 160734			
/complement(10775..12229) /note="(SAV2467" /codon_start=1 /transl_table=1 /product="putative secreted protein" /protein_id="BAC70179.1" /db_xref="GI:29606116"			241 -----IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAla 257			
CDS		complement(10775..12229) /note="(SAV2467" /codon_start=1 /transl_table=1 /product="putative secreted protein" /protein_id="BAC70179.1" /db_xref="GI:29606116"			Db	160733 TCCGCCAAGGTCCGCTCGCGCGCGGTACGACACACACCGCGAGCTGTACGTGACG 160674
	/translation="MRPRLPLFAARRGRSARRRTPSVLGAVGLATATATACDSGD DTKASSESSPASAGDDKIKIPDDIRDKLKERGIDIDKWKNGAWKNWRDMLREAN			258 ThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerVal 277		
	/complement(10775..12229) /note="(SAV2467" /codon_start=1 /transl_table=1 /product="putative secreted protein" /protein_id="BAC70179.1" /db_xref="GI:29606116"			160673 TACGCGCAGCACCCGCGCGCGCGCCAGTCCGCGCGAGCGTGCACAAAGTGCGTAC 160614		
	/translation="MAVTKSAVASLKREAGHAHAHEGILRRQSAESAARTYARALP IVPYRGLTTEGADRGVLDCLSGAGTLALGHNHPVVLAIKVLDSGAPLHVLDLA TPVDATFIELRPLAGLADRARVPCGPGATDAVEAAFKLVRATRGSLAFTGA YHMTAGALEASGATDGRVARLPYPQDYPFGIGGREGAELGARWTESLDDVKS PLPAGMLLEPVQEGGVIIPAPDAMLRMRATITADRSIPLIADSVQTVGRTGRFWAV EHSVPDVPVMSKGLPLAVVYRDDLDVWQPGAHAGTFRGNQLAMAAAGTATLA YVRHRLAERAQTLGARMLDOLSLATEFCPCVGEVRGRLMIGVELVRAEDRAVSP GSDPASLDTAGRPAAPALAAVQRECLRGILVELGGRHASVWRLPLPLTISDE QATAVLDKLDAVEARVARGRAERSGRFORSERAG"			278 ThrSerGlyThrTrpThrArgIleSerProVal-----ProSerThrAspThr 293		
	/complement(10775..12229) /note="(SAV2467" /codon_start=1 /transl_table=1 /product="putative secreted protein" /protein_id="BAC70179.1" /db_xref="GI:29606116"			160613 GCCACGGGACGTGGACCGAGGTACCCCGGTGAACCGCGCGGCGGACGAGCGACGCGC 160554		
	/translation="MRPRLPLFAARRGRSARRRTPSVLGAVGLATATATACDSGD DTKASSESSPASAGDDKIKIPDDIRDKLKERGIDIDKWKNGAWKNWRDMLREAN			294 AlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIle 313		
	/complement(10775..12229) /note="(SAV2467" /codon_start=1 /transl_table=1 /product="putative secreted protein" /protein_id="BAC70179.1" /db_xref="GI:29606116"			160553 TCGGCGGACACCTTCGCTACGCGGGGTGCGCGTGCAGCGCGCGCGCGCGCCCTC 160494		
	/translation="MAVTKSAVASLKREAGHAHAHEGILRRQSAESAARTYARALP IVPYRGLTTEGADRGVLDCLSGAGTLALGHNHPVVLAIKVLDSGAPLHVLDLA TPVDATFIELRPLAGLADRARVPCGPGATDAVEAAFKLVRATRGSLAFTGA YHMTAGALEASGATDGRVARLPYPQDYPFGIGGREGAELGARWTESLDDVKS PLPAGMLLEPVQEGGVIIPAPDAMLRMRATITADRSIPLIADSVQTVGRTGRFWAV EHSVPDVPVMSKGLPLAVVYRDDLDVWQPGAHAGTFRGNQLAMAAAGTATLA YVRHRLAERAQTLGARMLDOLSLATEFCPCVGEVRGRLMIGVELVRAEDRAVSP GSDPASLDTAGRPAAPALAAVQRECLRGILVELGGRHASVWRLPLPLTISDE QATAVLDKLDAVEARVARGRAERSGRFORSERAG"			314 MetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGly 333		
	/complement(10775..12229) /note="(SAV2467" /codon_start=1 /transl_table=1 /product="putative secreted protein" /protein_id="BAC70179.1" /db_xref="GI:29606116"			160493 GTCGTCTCCACCAACAACCGCTGGCGCGACGCGGACACCGTCTTCGCGTCCACGACGCGC 160434		
	Alignment Scores:	US-09-917-376-3 (1-740) x AP005031 (1-299550)		Qy Db Qy Db Qy Db Qy Db Qy Db		6 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 25
161381 TACCGCTGGGCGAACGCGGTACGGGGGACCGCGTTCGTCAACCGCGGTCTCTCCAC 161322			26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArgTrpAsp 45			
161321 CCCTCCGTACGCGGTCTCGGCTACGCCCGGACCGACATCGCGCGCGCTTACCGCTGGGAC 161262			46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr 65			
161261 GACCGCGCGCGCGCTCGACCCCGCTCATCGACCACTCGGCTGGGACGACGCTGGAACCTC 161202			66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85			



Db	160241	-----GCACCGCGGATCCGCGCTCGGAGGAGCGCGTGGCG	160203
Qy	434	AspLeuIleSerProSerGly---AlaProIleuSerAlaLeuGlyAspLeuGly	452
Db	160202	CAACTGATCTCGCCCGGTCGGGAGGACACCTGATCAGCGACTCGGGGACATCGGT	160143
Qy	453	GlyPheThrHisAlaAspValThrAlaValProSerThr---IlePheThrSerProVal	471
Db	160142	GTGATGTACACAGAGCGCTCAGCGGCTCGTGGCGGCATCGGACCAACCCGTG	160083
Qy	472	PheThrThrGlyThrSerValAspTyrAlaGluLeuAenProSerIleIleValargAla	491
Db	160082	TTGCGGTGCGGACGGACTCGCGCAGCGCGCGCGCGCGGTATGTGTGTCGACG	160023
Qy	492	GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly	511
Db	160022	GGCTGGGCGCACCGGACCGGC-----GGCTACTCCACGACGCG	159981
Qy	512	GlyLysAsnTyrP---PheGlnGlySerGluProGlyGlyValThrThrGlyGlyThr	529
Db	159980	GGCGGAGCTGGCGCCCTTCAGGCGCCAGCCCGACATCGCCAGGACGCGCGGACCG	159921
Qy	530	ValAlaAlaSerAlaAspGlySerArgPheValTrrpAla-----ProGlyAsp	545
Db	159920	ATCGCCACCATGTCGCGCGGCGGACACTGCTGTGTCTTCGTGCTGCGGACGCGC	159864
Qy	546	ProGlyGlnProValValTrrpAla-----ValGlyPheGlyAsnSerTrrpAla	561
Db	159863	-----ACGACGTACGCGGCCCGCCACCGCTCGACGACACCGCGCGAGCTGGTCC	159816
Qy	562	AlaSerGlnGlyValProAlaAsnNlaGlnIleArgSerAspArgValAsnProLysThr	581
Db	159815	GAGGTCCTCTCTCCGAGCGGCCACACCGCTCGCGGACCCCGCGCATCCGACGCGC	159756
Qy	582	PheTyrAla-----LeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThr	599
Db	159755	TTCTAGCTACGACTTCGACATGAGACGCTATAGCGCAGCAGCTGACAGTGGCGCTCG	159696
Qy	600	PheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAla	619
Db	159695	TTTCAGCGCGCGGCGGCGGACTGCC---TCCGCGCAGCAGCAGCTTCAAGTGTGTCGC	159639
Qy	620	ValProGlyLysGluGlyAspLeuTrrpLeuAlaAlaSer---SerGlyLeuTyrHisSer	638
Db	159638	GCGCGGAGCGGAGCGGACCTGTGTCTTCGCGCAATGCAACGCGGCTTACCGGTCC	159579
Qy	639	ThrAsnGlyGlySerSerTrrpSerAlaIleThrGlyValSerSerAlaValGly	658
Db	159578	ACCGAGCGGGGGAACCTTCGCGAGATCGACAGCTGCGGCTCGTACACCTCGCG	159519
Qy	659	PheGlyLysSerAlaProGlySerSerTyrProAlaValPheValGlyThrIleGly	678
Db	159518	TTTCGCGAGCGCGCGGCGCGGCGGCGGCGGATCTACCGTGGCTCGGCGGAG	159459
Qy	679	GlyValThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpValIleAsnAsp	698
Db	159458	ACCATCCCGCGCTTACCGCTCCGACGCGCGCGGACATCGGTTCGGATCAACGAC	159399
Qy	699	AspGlnHisGlnTyrGlyAsnTrrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArg	718
Db	159398	GACGCCACCATGTGGGGTGGATTCGCGGAGCGCGCTCGTGGTGACCGCGCATCCACGCG	159339
Qy	719	ArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAsp	733
Db	159338	CGGGTCTACTCGCCACCAACGCGCGCGCATCCAGTACGGGGAG	159294
RESULT 14			
AE011809			
LOCUS			
DEFINITION	Xanthomonas axonopodis pv. citri str. 306, linear BCT 29-MAY-2002		
ACCESSION	AE011809 AE008923		
VERSION	AE011809.1 GI:21107967		

KEYWORDS	Xanthomonas axonopodis pv. citri str. 306
SOURCE	Xanthomonas axonopodis pv. citri str. 306
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
REFERENCE	Xanthomonadales; Xanthomonas.
AUTHORS	1 (bases 1 to 14520) da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R., Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida Jr., N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C., Chamargo,F., Czapina,L.P., Cicarelli,R.M.B., Coutinho,L.L., Cursino-Santos,J.R., El-Dorri,H., Faria,J.B., Ferreira,A.J.S., Ferreira,R.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C., Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite Jr., R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A., Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J., Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M., Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira,Jr.,H.A., Rosa,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinola,L.A.F., Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos Santos,M., Truffi,D., Tsai,S.M., White,P.F., Setubal,J.C. and Kitajima,J.P.
TITLE	Comparison of the genomes of two Xanthomonas pathogens with differing host specificities
JOURNAL	Nature 417 (6887), 459-463 (2002)
MEDLINE	2202145
PUBMED	1202417
REFERENCE	2 (bases 1 to 14520) da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R., Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida Jr., N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C., Chamargo,F., Czapina,L.P., Cicarelli,R.M.B., Coutinho,L.L., Cursino-Santos,J.R., El-Dorri,H., Faria,J.B., Ferreira,A.J.S., Ferreira,R.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C., Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite Jr., R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A., Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J., Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M., Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira,Jr.,H.A., Rosa,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinola,L.A.F., Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos Santos,M., Truffi,D., Tsai,S.M., White,P.F., Setubal,J.C. and Kitajima,J.P.
TITLE	Direct Submission
JOURNAL	Submitted (28-NOV-2001) Departamento de Bioquimica, Universidade de Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900, Brazil
FEATURES	Location/Qualifiers 1..14520 /organism="Xanthomonas axonopodis pv. citri str. 306" /mol_type="genomic DNA" /strain="306" /db_xref="taxon:190486" /note="pathovar: citri" 87..2342 /gene="celA" /note="synonym: XAC1770" 87..2342 /gene="celA" /note="identified by sequence similarity; putative; ORF located using BlastX/Glimmer/Genemark" /codon_start=1 /transl_table=1 /protein_id="AA036634.1" /db_xref="GI:21107968" /translation="MIAVIDPPTKRFRSAHAWLTAVCLTLVLLFWTSEASDRPGTP DPYQHSVAIGGGFVTGLFHPAERDLAYATDVGAATKWDQAQOQWALTDLWLGAD NWLGMIDAFVADPADPNALYLAAGTYMHERAGNAVLRSFDRGKQFPAADLPKLG NOLGRANGLAVDPHGRVLLIGSDRGLMWSDDRGHAWRVEGFPADLAGATARN HYGSEAVGIAFVVPDAASHGAGATPRIYGVSTYATQSLTVSEDAGRWTSVAGQPK GLRPSHWGHNAGQWLYSGDRPGPDLMAGGALWNYDATQGRWREISPIQPATGDGF







```

Db 393 GATTGGAATCTGATGGGAATCGACGCTTTTGGGTGATCTCTGCCGACCCGGAACGCGTTG 452
Qy 82 TrrAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArg 101
Db 453 TACCTGGCTGTGGCACTACATCGACGAACGCGCCGCAAT---GCCGCGGTGTGGT 509
Qy 102 SerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLeuLeuGlyValAsn 121
Db 510 TCGTTTCGATCCGGTAAGCAGTTTGAACGCCGCCGACCTGCCGTTTCAGCTCGTGGCAAT 569
Qy 122 MetProGlyArgGlyMetGlyValGlyLeuAlaValAspProAsnAspAsnIleLeu 141
Db 570 CAATTGGCGCGCAATGGCAGCGCTGCGCTGACCGCGCACGACGCGCGGTGCTG 629
Qy 142 TyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrp 161
Db 630 TTGCTTGGCTCACGTGATGCG---GGCCTGTGGCGCACGACGATCGCGCGCGCACTGG 686
Qy 162 SerGlnMetThrAsnProAspValGlyThrTyrIleAlaAsnProThrAspThrThr 181
Db 687 GCGCGGTGGGAAGATTCTCTGCCGATGCAATTGGCTGGCGCCACGCGCGCAACCATGTC 746
Qy 182 GlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSer 201
Db 747 GGAAGCAACAGCGCGTG---GGCATCGCTTCGTGGTTCGACGCGCGCGCGCCAT 803
Qy 202 LeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPhe 221
Db 804 GCTGGCGCTCGACACACGACGATCTATGTGGCGTTTCGACTGGCGCAACACGACGCTGTAT 863
Qy 222 TrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIle 241
Db 864 GTCTCCGAAGATGGGGCGCACCTGTGTGGCGCTGGCGCGCGCAACCAAGGGCGTGGCG 923
Qy 242 ProHisLysGlyValPheAspProValAsnHisValLeu----- 254
Db 924 CCC-----AGCCACATGGTGGTCCGTCACACGCTGGCGGAGTGG 959
Qy 255 TyrIleAlaThrSerAsnThrGlyLysProTyrAspGlySerSerGlyAspValTrpLys 274
Db 960 TACTTGAGTATGGGATCGCCCGCGCGCGATCTGATGGCGGTGGCGCGCTGTGGAAC 1019
Qy 275 PheSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAspThrAla 294
Db 1020 TACGACGCGACGCAAGCGCGCTGGCGCGAGATCAGCCGATTTCCACACACGCGCACT--- 1076
Qy 295 AsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMet 314
Db 1077 GGCATGTGTTTGGCTGGGGCGCGGTGGCGGTGGATCCGCAACATCCGAGGTGCTGCTG 1136
Qy 315 ValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheAspSerThrAspGlyGly 334
Db 1137 GCCAGCAGTTCCGCGCGCGCACGCGCGCGATGAGTGTTCGAGCGCGGATGGCGGA 1196
Qy 335 AlaThrTrpThrArgIleTrpAspTrpSerTyrProAsnArgSerLeuArgTyrVal 354
Db 1197 GCGAGTGGTTCGCTGTGGCGCTGCGAGTTC----- 1232
Qy 355 LeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProProValPro 374
Db 1233 ---GATCACAGCGCGCGCGCTGG---ACCGCGCACGCAACACCGCAC 1274
Qy 375 SerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg 394
Db 1275 -----TGATAGCC---GGCTGGCGATCGATCCGTTTCATAGCAACAC 1316
Qy 395 MetLeuTyrGlyThrGlyAlaThrLeuTyrTrpAlaThrAsnAspLeuThrLysTrpAspSer 414
Db 1317 GCCAGTTTCGTCACCGCGCTACGGCATCGGGCTCGCGCAATCTCGACAGCTTTTCAGAG 1376
Qy 415 GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAsp 434

```

```

Db 1377 CAGCAGCGCGTGCAGTGTGGTTCAGAGCAGCGTGGCGTGGAGAAACCGTGGCGCTGGAT 1436
Qy 435 LeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPhe 454
Db 1437 CTGCTCAGCCGATGCCCGCGCGGATCTGCTCAGCGCGCTGGCGGATATCGATGGCTTC 1496
Qy 455 ThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThr 474
Db 1497 CGGCAT---GACCGCTGGACACCGCACAGTTCAGTATCTTGGTCCGCGCTGACCAAT 1553
Qy 475 GlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValargAlaGlySerPhe 494
Db 1554 GCGCAAGACATCGATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1613
Qy 495 AspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsn 514
Db 1614 ---CCGCGATCGTGGCAACACGAGATTCGCGCGCTGTATTCGAGGACGGGGGACGCGAC 1670
Qy 515 TrpPheGln---GlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSer 533
Db 1671 TGGGCTGCATTCGCCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1730
Qy 534 AlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAla 553
Db 1731 GCCGATGCATCGCAGGTGTGTGGTGGCGGACCAAGCGCGG-----GTGTGGCGC 1781
Qy 554 ValGly---PheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIle 572
Db 1782 ACTGCGCATTCGCCAAGCGTGGCAGCGCTGCAGGCGCTTGCCTGATACCGCGGTGGT 1841
Qy 573 ArgSerAspArgValAsnProLysThrPheTyrAlaLeuSer-----AsnGlyThrPhe 590
Db 1842 GTGGCGGACCGGTGATCGCGCAGCGCTGGTATCGCGCGCGCGCGCGCGCGCGCGCGCTG 1901
Qy 591 TyrArgSerThrAspGlyGlyValThrPheGlnProValAla-----AlaGlyLeuPro 608
Db 1902 TACGAGAGCAGCATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1961
Qy 609 SerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysGlyAspLeuTrp 628
Db 1962 GCGCGCGATGAA-----CGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2006
Qy 629 LeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSer----- 646
Db 2007 CGTGGCGCGGTGTGTATCTGGCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2066
Qy 647 -----AlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAla 663
Db 2067 CAGTTGCGCACGCTATCCAAATCAGATGAAGCGCGCTCGCTGGGCATCGCGAAGCGCTA 2126
Qy 664 ProGlySerSerTyrProAlaValPheValGlyThrIleGlyGlyValThrGlyAla 683
Db 2127 CGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2186
Qy 684 TyrArgSerAspAspCysGlyThrTrpValLeuIleAsnAspAspGlnHisGlnTyr 703
Db 2187 TTCCGCTCCGACGACGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2246
Qy 704 GlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgValTyrIleGly 723
Db 2247 GGCNAA---CCGTATAGCGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2303
Qy 724 ThrAsnGlyArgGlyIleValTyrGlyAsp 733
Db 2304 ACCGCGCGCGGTGGCATCTTCTATGGCAT 2333

```

RESULT 15  
 AE001712  
 LOCUS  
 DEFINITION  
 Thermotoga maritima MSB8 section 24 of 136 of the complete genome.  
 AE001712  
 ACCESSION  
 AE001712.1 GI:4980799  
 VERSION  
 KEYWORDS







NSKFAVHTYTPFLSYSDGGINWREIETVPEGITDGGRLSLAVNSDGGKTLVWSPA  
 NHEVIVSDKGSKWKALSVPEFENYPPASDPVNPSPKEVIFDWNKGDFLISXDDGGS  
 FMKAGTDPKRWVSLYSPFLAPDRGDIWALQWNGLYRSKGGITFERLGNVDI  
 AVYVFGAPKPGDYDPAIYLVNPGVGVIFMSTDEGKTWNRINNDKHQFGWIHYMIG  
 DWNFGRIFLGTGRGRIIVGEVKEE"  
 8014..9363

gene

/locus\_tag="TM0306"

CDS

/locus\_tag="TM0306"

/product="alpha-L-fucosidase, putative"

/protein\_id="AAD35394.1"

/db\_xref="GI:4980806"

/transl\_table=11

/product="alpha-L-fucosidase, putative"

/protein\_id="AAD35394.1"

/db\_xref="GI:4980806"

/transl\_table=11

/product="alpha-L-fucosidase, putative"

/protein\_id="AAD35394.1"

/db\_xref="GI:4980806"

/transl\_table=11

/product="alpha-L-fucosidase, putative"

/protein\_id="AAD35394.1"

/db\_xref="GI:4980806"

/transl\_table=11

/product="alpha-L-fucosidase, putative"

/protein\_id="AAD35394.1"

/db\_xref="GI:4980806"

/transl\_table=11

/product="alpha-L-fucosidase, putative"

/protein\_id="AAD35394.1"

/db\_xref="GI:4980806"

/transl\_table=11

/product="alpha-L-fucosidase, putative"

/protein\_id="AAD35394.1"

/db\_xref="GI:4980806"

/transl\_table=11

/product="alpha-L-fucosidase, putative"

/protein\_id="AAD35394.1"

/db\_xref="GI:4980806"

/transl\_table=11

/product="alpha-L-fucosidase, putative"

/protein\_id="AAD35394.1"

/db\_xref="GI:4980806"

/transl\_table=11

/product="alpha-L-fucosidase, putative"

/protein\_id="AAD35394.1"

/db\_xref="GI:4980806"

/transl\_table=11

/product="alpha-L-fucosidase, putative"

/protein\_id="AAD35394.1"

/db\_xref="GI:4980806"

/transl\_table=11

/product="alpha-L-fucosidase, putative"

/protein\_id="AAD35394.1"

/db\_xref="GI:4980806"

/transl\_table=11

/product="alpha-L-fucosidase, putative"

/protein\_id="AAD35394.1"

/db\_xref="GI:4980806"

/transl\_table=11

/product="alpha-L-fucosidase, putative"

/protein\_id="AAD35394.1"

/db\_xref="GI:4980806"

/transl\_table=11

/product="alpha-L-fucosidase, putative"

/protein\_id="AAD35394.1"

/db\_xref="GI:4980806"

/transl\_table=11

/product="alpha-L-fucosidase, putative"

/protein\_id="AAD35394.1"

/db\_xref="GI:4980806"

/transl\_table=11

/product="alpha-L-fucosidase, putative"

/protein\_id="AAD35394.1"

/db\_xref="GI:4980806"

/transl\_table=11

/product="alpha-L-fucosidase, putative"

/protein\_id="AAD35394.1"

/db\_xref="GI:4980806"

/transl\_table=11

6134 ATGGAGTGTTCAGCGTGTCTCTCGATCTCGAGAGAAATCTACGCGATGACA 6193  
 86 GlyMetTyrThrAsnSerTrpAspProAsnAspGlyValAlaIleLeuArgSerSerAspGln 105  
 6194 GGAATAACATCAGTGGATTTGGAGGATTTCTTATCTCGAGATATAC 6250  
 106 GlyAlaThrTrpGlnIleThrProLeuPro-----PheLysLeuGlyGlyAsnMet 122  
 6251 GGTGAACGTTGGACAACTTGAACCTTGCAAGATATGGATATAAGTTGGTGGACGAA 6310  
 123 ProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyr 142  
 6311 GATGAAGAAACGCTGTTGAAAGATTGCAAGTAGATCCAAATTTTCAGTAGCGTCTTTT 6370  
 143 PheGlyAlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThrTrpSer 162  
 6371 ATGGGTACA--ACGAAATATGCTCTGTGGAATCAGAGATTTTGGAAAAAATTTGAAA 6427  
 163 GlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGly 182  
 6428 AAGTGGATAGTTTTCCCTCAACC-----AGTCTAACGTTTGTCTATTCGATGAGAAAAAGTGGAAAAA 6493  
 183 TyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeu 202  
 6494 GGAAGTCTCTACTCTAGAAATTTTGTGGATGTTTCAGAACCCAGGGG--ATCTTTGTA 6550  
 223 SerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIlePro 242  
 6551 ACTGAAGATGGTGGAAACCACTGGAAATGTTCTTCCAAACCTGCCGAAAGCATCTCATTC 6610  
 243 HisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGly 262  
 6611 TTGGTGGAAAAATACAT-----GACGGGATCTTTACGTTACACATTTCCATGCTTTG 6664  
 263 GlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrp 282  
 6665 GGACCGAATGGTGTCTACGAGAGGGGCTGTTATGAAGTATGATGATGCTGATCAAAAAATG 6724  
 283 ThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGly 302  
 6725 TACGATGTAAACCAATGAAGGAGAT-----TTCCGATATTTGTTGA 6766  
 303 LeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrp 322  
 6767 ATAGATGTTCAAGAG-----AACGTGGTGTATGATATCCACATCGATAGTGGTAT 6817  
 323 ProAspThrIleIlePheArgSerThrAspGlyGlyValAlaThrTrpThrArgIleTrpAsp 342  
 6818 CCACACGATGAGATATTTATTTCCCTGACGAGGAGAAACGTGGAGACCTCTTCTGGAA 6877  
 343 TrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrp 362  
 6878 AAGGCAAAATTT-----GATATCAACAAACGCGCCCTGG 6910  
 363 LeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTrpMetAsp 382  
 6911 ATAAAA-----GATTAATCTCT-----CACTGGATATCT 6940  
 383 GluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThr 402  
 6941 GAC--GTCAGATAGATCCATTTGACATGAATAGGCTATTTTACACACAGGTATGCT 6997  
 403 LeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaPro 422  
 6998 GTGTGGGTGCTTATGAGTTAAAAAAGTCAATCGAAGGT-----ATGGGAAA 7045  
 423 MetValLys-----GlyLeuGluGluThrAlaValAsnAspLeuIle 436

Alignment Scores:

Pred. No.:	7,72e-50	Length:	13278
Score:	1160.50	Matches:	264
Percent Similarity:	51.53%	Conservative:	123
Best Local Similarity:	35.15%	Mismatches:	275
Query Match:	28.75%	Indels:	89
DB:	1	Gaps:	26

US-09-917-376-3 (1-740) x AB001712 (1-13278)

6 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 25  
 5954 TTTGAGTGGAAATCGTGGAAATCAACGAGGAGGATTTGTACAGGAATAATATTCAT 6013  
 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAsp 45  
 6014 CCGTAGTCTCGGCTTTTATATCGGAGAACGAGTGTGGTGGATTTTACAGGTGGAT 6073  
 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTyrAsnAsnTrpGlyTyr 65  
 6074 GAGGAACAAAGAGTGGAGCAATTTTGTATTTTGTAGNAGAGATCAGTCCGATATAC 6133  
 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85



Db 7046 CCAGTGAATCGATATTTCGAAATCGAGGGTTAGAGAAAACAGTTGTTTGCACACTGGTT 7105  
Qy SerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyPheThrHis 456  
Db 7106 CCCCCCATAGAGAAAGCCCTCTCTCTGCAATCGCGATTCGGGAGGATCCGTCAC 7165  
Qy 457 AlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThr 476  
Db 7166 GAAAGTTTGGACACCCACCACATCATCTATGTAC---AAACCCCTAAATGGACGCTCTCTT 7222  
Qy 477 SerValAspTyrAlaGluLeuAsnProSerIleIleValAlaGlySerPheAspPro 496  
Db 7223 GGTATAGCTTTTGCTATCAGAAATTCAAATTCGTGGCAGAGTGCATACATAC------ 7276  
Qy 497 SerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPhe 516  
Db 7277 ---ACCTATCTCT-----TTCCTGTCTTACTCCGAGGATGGCGAATCAACTGGAGA 7324  
Qy 517 Gln---GlySerGluProGlyGlyValThrThrGlyGly-----ThrValAlaAlaSer 533  
Db 7325 GAGATAGAAACTGTCGGAGGGAATACTGACGCGGGAGGTTAAGTTTGGCAGTCTCG 7384  
Qy 534 AlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValThrAla 553  
Db 7385 AACGATGGAATAACACTGGTCTGGAGTCCAGCGAAT-----CATGAGGTAATTTGTTC 7438  
Qy 554 ValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAla-----AsnAla 570  
Db 7439 AGCGATAAAGGAAAGCTGGAAAGCCATTAGTTGTCAGTTCCAGATTCCAACTAT 7498  
Qy 571 GlnIleArgSerAspArgValAsnProLysThrPheTyrAlaLeu-----SerAsnGly 588  
Db 7499 TTTCTCTCATCTGATCTGTAAATCCGAGTAAATTTTACATCTTCGATTGGAAAAACGGT 7558  
Qy 589 ThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuPro 608  
Db 7559 GACTTCTCTATAAGTAAAGATGGAGGAAGAGTTTATGAAA---GGTGCTAAACTTCCT 7615  
Qy 609 SerSerGlyAlaValGlyValMetPheHisAla-----ValProGlyLysGlu 624  
Db 7616 TCTTTTGATAACTGGTGGGTATCGCTTTATTTCTTCCCTGTGTGGCTCTCTGACAGAGAA 7675  
Qy 625 GlyAspLeuTrpLeuAlaAlaSer---SerGlyLeuTyrHisSerThrAsnGlyGlySer 643  
Db 7676 GGAGATATCTGTTAGCATTCAATGGAATGGACTTTACAGATCAAAAGATGGTGAATA 7735  
Qy 644 SerTrpSerAlaIleThrGlyValSerSerAlaValAlaAsnValGlyPheGlyLysSerAla 663  
Db 7736 ACGTTTGAGGACTAGGAAACGTTGATATAGCTACGTGATTGGTTTCGGAGCCCAAAA 7795  
Qy 664 ProGlySerSerTyrProAlaValPheValValGlyThrIleGlyGlyValThrGlyAla 683  
Db 7796 CCAGGTACTGACTATCCAGCAATTTACCTGAACGGTATGGTGAATGGAGTTTATGGTATT 7855  
Qy 684 TyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAspGlnHisGlnTyr 703  
Db 7856 TTCATGTCGACAGATAGGGGCAACATGGATGAGAAATTAACACAGATAGCATCAATTT 7915  
Qy 704 GlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArgValThrIleGly 723  
Db 7916 GGA---TGGATACATTACATGATAGAGATATGAATGAATTCGGGAGAAATTTCTCTGGT 7972  
Qy 724 ThrAsnGlyArgGlyIleValTyrGlyAspIle 734  
Db 7973 ACGGAAGGAGAGGAAATCATCGTTGGTGAAGTG 8005

Search completed: October 5, 2005, 16:33:04  
Job time : 10380.6 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model  
Run on: October 5, 2005, 06:17:10 ; Search time 1138.43 Seconds  
(without alignments)  
3847.935 Million cell updates/sec

Title: US-09-917-376-3  
Perfect score: 4036  
Sequence: 1 ATPTQYTSWNAICGGGFGVD.....YIGTNGRGIVYGDIGGAPSG 740

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters: -DB=xlh  
-MODE=frame+ p2n.model -DEV=xlh  
-Q/cn2\_1/USPTO spool/US09917376/runat\_04102005.164327.26889/app\_query.fasta\_1.2524  
-DB=N Geneseq 16Dec04 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MAIRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09917376 @CN 1.1 1867 @runat\_04102005.164327.26889 -NCPU=6 -ICPU=3  
-NO MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 16Dec04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004as:\*  
13: geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4036	100.0	2869	10 ADD22922	Add22922 Acidother
2	4036	100.0	2869	10 ABZ77632	Abz77632 Nucleotid
3	4036	100.0	2869	12 AD052314	Ado52314 Acidother
4	2063	51.1	2823	8 ABV76941	Abv76941 Nucleotid
5	1579	39.1	2517	12 ADH51583	Adh51583 T reesei

6	1579	39.1	2710	12 ADH51580	Adh51580 Trichoder
7	1506	37.3	2849	10 ADD42060	Add42060 Trichoder
8	1139	28.2	3668	10 ADD24893	Add24893 DNA encod
9	1117	27.7	2367	12 ADH19119	Adh19119 Geotrichu
10	1117	27.7	2646	12 ADH19117	Adh19117 Geotrichu
11	1114	27.6	2481	12 ADH19123	Adh19123 Geotrichu
12	979	24.3	2268	13 ADR90297	Adr90297 Geotrichu
13	979	24.3	2274	13 ADR90301	Adr90301 Geotrichu
14	979	24.3	2755	13 ADR90295	Adr90295 Geotrichu
15	878.5	21.8	5698	10 ADD42054	Add42054 Trichoder
16	392.5	9.7	1103	3 AAF14988	Aaf14988 Trichoder
17	238	5.9	7407	8 ACA42281	Aca42281 Prokaryot
18	238	5.9	7407	9 ACC59398	Acc59398 Microbial
c 19	227.5	5.6	3129	11 ABD15269	Abd15269 Pseudomon
c 20	227.5	5.6	8211	11 ABD15052	Abd15052 Pseudomon
c 21	203	5.0	4188	11 ABD15170	Abd15170 Pseudomon
c 22	201.5	5.0	4131	11 ABD15169	Abd15169 Pseudomon
c 23	199.5	4.9	7305	8 ACA23399	Aca23399 Prokaryot
c 24	195	4.8	5295	8 ACA24001	Aca24001 Prokaryot
c 25	192	4.8	9903	8 ACA40315	Aca40315 Prokaryot
c 26	192	4.8	110000	4 AAI99682_04	Continuation (5 of
c 27	188.5	4.7	4399	13 ADS61968	Ads61968 Bacterial
c 28	184	4.6	3324	8 ACA38418	Aca38418 Prokaryot
c 29	184	4.6	110000	4 AAI99683_04	Continuation (5 of
c 30	182	4.5	11148	8 ACF39394	Acf39394 Mycobacte
c 31	182	4.5	69350	12 ADM98959	Adm98959 Diterpene
c 32	182	4.5	110000	4 AAI99682_37	Continuation (38 o
c 33	182	4.5	110000	4 AAI99683_37	Continuation (38 o
c 34	181.5	4.5	11679	9 ADA30090	Ada30090 DNA encod
c 35	180	4.5	2355	13 ADS62118	Ads62118 Bacterial
c 36	179.5	4.4	3222	8 ACA26562	Aca26562 Prokaryot
c 37	177.5	4.4	110000	4 AAI99682_30	Continuation (31 o
c 38	177.5	4.4	110000	4 AAI99683_30	Continuation (31 o
c 39	176.5	4.4	9855	10 ACF04821	Acf04821 M.licheni
c 40	176.5	4.4	51855	10 ACF04818	Acf04818 Melithiaz
c 41	176	4.4	32367	2 AAV35620	Aav35620 Human SHO
c 42	174.5	4.3	67459	8 ABQ77491	Abq77491 S. aurant
c 43	174	4.3	2855	4 ABL29005	AbL29005 Drosophil
c 44	174	4.3	5221	4 ABL19628	AbL19628 Drosophil
c 45	174	4.3	5302	4 ABL29004	AbL29004 Drosophil

ALIGNMENTS

RESULT 1  
ADD22922  
ID ADD22922 standard; DNA; 2869 BP.  
XX  
AC ADD22922;  
XX  
DT 15-JAN-2004 (first entry)  
XX  
DE Acidothermus cellulolyticus avicelase Aviii DNA.  
XX  
KW Aviii; cellulose reduction; agricultural biomass; municipal solid waste;  
glycoside hydrolase; avicelase; ds; gene.  
XX  
OS Acidothermus cellulolyticus.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..2868  
FT /\*tag= a  
FT /product= "Aviii"  
XX  
PN US2003108988-A1.  
XX  
PD 12-JUN-2003.  
XX  
PF 18-OCT-2002; 2002US-00155400.  
XX  
PR 28-JUL-2001; 2001US-00917376.  
XX  
PA (DING/) DING S.



PA (ADNE//) ADNEY W S.  
 PA (VINZ//) VINZANT T B.  
 PA (HIMM//) HIMMEL M E.  
 XX  
 PI Ding S, Adney WS, Vinzant TB, Himmel ME;  
 XX  
 XX WPI; 2003-810853/76.  
 DR P-PSDB; ADD22921.  
 XX  
 PT New isolated thermal tolerant avicelase polynucleotide useful for  
 PT detection of a polynucleotide encoding AviiiI and for reducing cellulose  
 PT in a starting material, e.g. municipal solid waste.  
 XX  
 XX Claim 17; SEQ ID NO 2; 29pp; English.  
 XX  
 CC The invention relates to an isolated polynucleotide molecule encoding a  
 CC thermostable AviiiI polypeptide. The polynucleotide is useful for  
 CC detection of a polynucleotide encoding AviiiI. The polynucleotide is  
 CC useful for reducing cellulose in a starting material which involves  
 CC administering to the starting material, e.g. agricultural biomass or  
 CC municipal solid waste, a polypeptide molecule of the polynucleotide. The  
 CC method further comprises administering a second polypeptide molecule  
 CC chosen from the glycoside hydrolase family of proteins. The present  
 CC sequence represents DNA encoding the Acidothermus cellulolyticus  
 CC avicelase AviiiI.  
 XX  
 SQ Sequence 2869 BP; 546 A; 857 C; 900 G; 565 T; 0 U; 1 Other;

Alignment Scores:  
 Pred. No.: 1,17e-251 Length: 2869  
 Score: 4036.00 Matches: 740  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: Gaps: 0

US-09-917-376-3 (1-740) x ADD22922 (1-2869)

QY 1 AlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAsp 20  
 DB 139 GCGACGACTACGCGGTACACCTGGAGCAACGTGGCGGATCGGGGGCGGCGCTTCGAC 198  
 QY 21 GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGly 40  
 DB 199 GGGATCGCTTCAATGAAGGTGCACCGGAATTCGTACGTGGCAGCAGCATCGGGGG 258  
 QY 41 MetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrp 60  
 DB 259 ATGTATCGATGGGATGCCCAACCGCGCGGTGGATCCCTCTTCTGGATTGGGTGGATGG 318  
 QY 61 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 80  
 DB 319 AACAAATGGGGGTACAAACGGCGTCTCAGCATTTGGCGCAGACCCCGATCAATACTAACAA 378  
 QY 81 ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 100  
 DB 379 GTATGGGCGCGCTCGGAATGTACACCAACAGCTGGGACCCAAACGACGAGCGATTCTC 438  
 QY 101 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGly 120  
 DB 439 CGCTCGTCTGATCAGGGCGCAACGTGGCAATAAACGCCCTCGCTTCAAGCTTGGCGGC 498  
 QY 121 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 140  
 DB 499 AACATGCCCGCGGTGGAATGGCGAGCGGCTTGGCGTGGATCCAAACAAATGACAACTT 558  
 QY 141 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThr 160  
 DB 559 CTGTATTTTCGGCCCGCGAGCGGAAAGGGCTCTGGAGAAGCAGACAGATTCGCGCGCGACC 618  
 QY 161 TrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 180  
 DB 619 TGGTCCCAGATGACCAACTTTTCGGACGATGAGCAGTACATTTGCRAAATCCCACTGACAG 678

QY 181 ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSer 200  
 DB 679 ACCGGCTATCAGAGCGATATTCAAGGCGTCTGGTCTGGTCTTCGACAAGTCTTGCTCA 738  
 QY 201 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal 220  
 DB 739 TCGCTCGGGCAAGCGAGTAAGACCATTTTGTGGCGTGGCGGATCCCAATAATATCGGTC 798  
 QY 221 PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyValaProThrGlyPhe 240  
 DB 799 TTCTGGAGCAGAGCGGCGCGAGCTGGAGGCGTGGCGGTGGCGGCGCGGCGGCTTC 858  
 QY 241 IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn 260  
 DB 859 ATCCCGCACAGGGGCTTTTGACCCGGTCAACACCGTCTCTATATGGCCACCAAGCAAT 918  
 QY 261 ThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGly 280  
 DB 919 ACGGGTGGTCCGTATGACGGGAGCTCCGGCGACGCTCTGGAATTTCTCGGTGACCTCCGG 978  
 QY 281 ThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr 300  
 DB 979 ACATGGACGCGAATCAGCCCGTACCTTCGACGGACACGGCCCAACGACTACTTTGGTTAC 1038  
 QY 301 SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer 320  
 DB 1039 ACGCGCTCACTATTCAGCCCGCACCCGCAACACGATATGTTGGCAACCCAGATATCG 1098  
 QY 321 TrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIle 340  
 DB 1099 TGGTGGCGGACACCATATCTTTGGAGCACCGACCGCGGTGCGAGCTGGACGCGGATC 1158  
 QY 341 TrpAspTrpThrSerTyrProAsnArgSerLeuLeuArgTyrValLeuAspIleSerAlaGlu 360  
 DB 1159 TGGGATTGGACGAGTTATCCCAATCGAAGCTTTCGATATGTGTTCGATTTTCGGCGGAG 1218  
 QY 361 ProTrpLeuThrPheGlyValGlnProAsnProValProSerProLysLeuGlyTrp 380  
 DB 1219 CTTTGGCTGACCTTTCGGGTACAGCCCAATCTCCCGTACCCAGTCCGAGCTCGGCTGG 1278  
 QY 381 MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGly 400  
 DB 1279 ATGGATGAAGCGATGCAATCGATCGTTCAACTCTGTATCGGATGTCTACGGAACAGGC 1338  
 QY 401 AlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIle 420  
 DB 1339 GCGACGTTGTACGCAACAAATGATCTCAGAAAGTGGGACTCCGGCGGCCAGATTCAATC 1398  
 QY 421 AlaProMetValLysGlyLeuGluThrAlaValAsnAspLeulleSerProProSer 440  
 DB 1399 GGGCGGATGGTCAAGAGATTGGAGGAGACGGCGGTAAACGATCTCATACCCCGCGCTCT 1458  
 QY 441 GlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyPheThrHisAlaAspValThr 460  
 DB 1459 GGGCGCCCGCTCATCAGCGCTCTCGGAGACTTCGGCGGCTTCACCCACGCGGAGTTACT 1518  
 QY 461 AlaValProSerThrIlePheThrSerProValPheThrThrThrGlyThrSerValAspTyr 480  
 DB 1519 GCGGTGCCATCGACGATCTTCAGTCCGGGTTCACGACCGCGCACCGGCTCGACTAT 1578  
 QY 481 AlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSerGlnPro 500  
 DB 1579 GGGGAATTGAATTCGCTCGATCATCTGCTCGCTGGAGATTTCGATCATCGAGCAACCG 1638  
 QY 501 AsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGlu 520  
 DB 1639 AACGACGACGCTCGCTTCGACAGAGCGCGGCAAGNACTGGTTCCAGGCGAGCGAA 1698  
 QY 521 ProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheVal 540  
 DB 1699 CCTGGCGGGGTGACGACGCGCGGACCGTCCGCGCATCGCGCGAGGCTCTCGTTTCTGTC 1758







```
Qy 181 ThrGlyTyrGlnSerAspIleGlnGlyValValTyrValAlaPheAspLysSerSerSer 200
Db 679 ACCGGCTATCAGACGATATTCAGGGCTGCTGGGGTCCGCTTCGACAAAGTCTTCGTCA 738
Qy 201 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal 220
Db 739 TCGCTCGGGCAAGCGAGTAAAGACCATTTTGTGGCGTGCGGGATCCCAATAATCCGGTC 798
Qy 221 PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 240
Db 799 TTCTGGAGCAGACGCGCGCGACGTGGCGCGGTGGCGCGACCGGGCTTC 858
Qy 241 IleProHsiLysGlyValPheAspProValAsnHsiValLeuTyrIleAlaThrSerAsn 260
Db 859 ATCCCGCACAAGGCGTCTTTGACCCCGTCAACACAGTGTCTATATATGACACAGCAAT 918
Qy 261 ThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGly 280
Db 919 ACGGTCGTCGTATGACGGAGCTCCGGCGAGCTCTGGAAATTTCTCGGTGACCTCCGGG 978
Qy 281 ThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr 300
Db 979 ACATGGACGGCAATCAGCCGGTACCTTCGACGACACGGCCACGACTACTTTTGGTTAC 1038
Qy 301 SerGlyLeuThrIleAspArgGlnHsiProAsnThrIleMetValAlaThrGlnIleSer 320
Db 1039 AGCGGCTCACTATCGACGGCCAGCACCCGAACAGATATATGTTGGGCAACCCAGATATCG 1098
Qy 321 TrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgile 340
Db 1099 TGGTGGCGGACACCAATAATCTTTTCGGAGCACCGCGCGGTGGAGCTGGACGCGGATC 1158
Qy 341 TrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGlu 360
Db 1159 TGGGATTTGGACGAGTATCCCAATCGAAGCTTGGGATATGTGTCATTTTCGGCGGAG 1218
Qy 361 ProTrpLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTrp 380
Db 1219 CCTTGGCTGACCTTTCGGGGTACACCGCAATCCTCCCGTACCAGTCCGAAGCTCGGCTGG 1278
Qy 381 MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGly 400
Db 1279 ATGGATGAGCGATGGCAATCGATCCGTTCAACTCTGATCGGATGCTCTACGGAAACAGGC 1338
Qy 401 AlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHsiIle 420
Db 1339 GCGACGTTGTACGCAACAAATGATCTCAGAAAGTGGGACTCCGGCGGCCAGATTCATATC 1398
Qy 421 AlaProMetValLysGlyLeuGluGluThrAlaValAsnAspLeuIleSerProSer 440
Db 1399 GCGCCGATGGTCAAGAGATTGGAGACGCGCGTAAACGATCTCATCAGCCCGCGCT 1458
Qy 441 GlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHsiAlaAspValThr 460
Db 1459 GCGCCCGCTCATCAGGCTCTCGGAGACCTCGGCGGCTTCACCCAGCCGCGAGTTACT 1518
Qy 461 AlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyr 480
Db 1519 GCGGTGCATCGACGATCTTCACGCTCAGCGTGTTCAGCACCGCACCGAGTTCGACTAT 1578
Qy 481 AlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSerGlnPro 500
Db 1579 GCGGAATTTGAATCCGTCGATCATGCTTCGCGCTGGAAATTTTCGATCCATCAGCAACCG 1638
Qy 501 AsnAspArgHsiValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGlu 520
Db 1639 AACGACAGGACGTCGCGCTTCTCGACAGACGCGCGCAAGAACTGTTCCAAAGGACGCGAA 1698
Qy 521 ProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheVal 540
Db 1699 CCTGGCGGGTGACGACGCGGGCGCACCGTCCGCGCATCGCGCGAGCTCTCGTTCTGTC 1758
```

```
Qy 541 TrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrp 560
Db 1759 TGGGCTCCCGCGATCCGGTCAGCTGTGTGTGTACGCGATCGGATTTGGCAACTCTCTGG 1818
Qy 561 AlaAspSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLys 580
Db 1819 GCTGCTTCGCAAGGTGTTCCGCCAATCCAGATCCGCTCAGACCGGGTGTAATCAAAG 1878
Qy 581 ThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPhe 600
Db 1879 ACTTCTATGCCCTATCCAAATGGAACTTCTATCGAAGCACGACGCGGGTGACATTC 1938
Qy 601 GlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHsiAlaVal 620
Db 1939 CAACCGGTTCGCGCGCGGTCTTCCGAGCAGCGGTGCGGTGTCTATGTTCCACGCGGTG 1998
Qy 621 ProGlyLysGluGlyAspLeuTrpLeuAlaIleAspSerGlyLeuTyrHsiSerThrAsn 640
Db 1999 CCTGAAAAGAAAGCGATCTGTGGCTCGCTGCATCGACGCGGCTTTACCACTCAACCAAT 2058
Qy 641 GlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGly 660
Db 2059 GCGCGCAGAGTGTGTCTGCAATCACCGCGTATCTCCGCGTGAAAGTGGGATTTGGT 2118
Qy 661 LysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGlyVal 680
Db 2119 AAGTCTGCGCGCGGTGCTCATACCCAGCGCTTTTGTGTCGCGCACGATCGAGGCGTT 2178
Qy 681 ThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAspAspGln 700
Db 2179 ACGGGCGGTACCGCTCCGACGACTGTGGGACGACCTGGGTACTCATCAATGATGCCAG 2238
Qy 701 HisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgVal 720
Db 2239 CACCAATACGGAATTTGGGACAAGCAATCACCGTGACCAACCGGAATTTAGCGGGTG 2298
Qy 721 TyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyGlyAlaProSerGly 740
Db 2299 TACATAGCACGACGCGCGTGGAAATTTGATATACGGGACATTTGTTGGTGGCGGTCCGGA 2358
RESULT 3
ADOS2314
ID ADOS2314 standard; DNA; 2869 BP.
XX
AC ADOS2314;
DT 15-JUL-2004 (first entry)
XX
DE Acidothermus cellulolyticus avicelase III (AviIII) DNA.
XX
KW Thermostable cellulase; avicelase III; AviIII; cellulose reduction;
KW biomass degradation; ethanol formation; industrial chemical;
KW fabric treatment; gene; ds.
XX
OS Acidothermus cellulolyticus.
XX
FH Location/Qualifiers
FT 1. .2868
FT /*tag= b
FT /product= "Acidothermus cellulolyticus avicelase III"
FT /note= "CDS contains translational exceptions"
FT /partial
FT /note= "No stop codon"
FT sig_peptide 1. .306
FT /*tag= a
FT mat_peptide 307. .2865
FT /*tag= c
FT /product= "Acidothermus cellulolyticus mature avicelase
FT III"
FT 2869
FT misc_feature /*tag= d
FT /note= "A string of unknown nucleic acid units"
XX
```



```
PN US2004038334-A1.
XX
PD 26-FEB-2004.
XX
PF 28-JUL-2001; 2001US-00917376.
XX
PR 28-JUL-2001; 2001US-00917376.
XX
PA (DING/) DING S.
PA (ADNE/) ADNEY W S.
PA (VINZ/) VINZANT T B.
PA (HIMM/) HIMMEL M E.
XX
XX Ding S, Adney WS, Vinzant TB, Himmel ME;
XX
XX WPI; 2004-203224/19.
DR P-PSDB; AD052313.
XX
PT Novel thermostable AviiiI polypeptide of glycoside hydrolase family and
PT isolated from Acidothermus cellulolyticus, useful for degrading treated
PT biomass into simpler forms of carbohydrate.
XX
XX Claim 17; SEQ ID NO 2; 19pp; English.
XX
XX The invention relates to a thermostable cellulase enzyme, avicelase III
XX (AviiiI) and its nucleic acid sequence. AviiiI is useful for reducing
XX cellulose in a starting material. A thermostable AviiiI peptide is useful
XX for degrading treated biomass into simpler forms of carbohydrate, which
XX is used in the formation of ethanol or other industrial chemicals. It is
XX also useful for treating fabrics to remove cellulose-containing stains.
XX The present sequence is Acidothermus cellulolyticus AviiiI DNA.
XX
XX SQ Sequence 2869 BP; 546 A; 857 C; 899 G; 565 T; 0 U; 2 Other;

Alignment Scores:
Pred. No.: 1,17e-251 Length: 2869
Score: 4036.00 Matches: 740
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 12 Gaps: 0

US-09-917-376-3 (1-740) x AD052314 (1-2869)

QY 1 AlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAsp 20
DB 139 GCGAGCACTCACCGGTACACTGGAGCAACGTGGCGATCGGGGGCGGCGCTTGTGCAC 198
QY 21 GlyileValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGly 40
DB 199 GGGATCGTCTTCAATGAAGTGCACCGGGAATCTGTAGTGGGACGACATCGGGGG 258
QY 41 MetTyrArgTTrpAspAlaAlaAsnGlyArgTTrpIleProLeuLeuAspTrpValGlyTTrp 60
DB 259 ATGTATCGATGGATGGCGCCCAACGGCGGTGATCCCTCTTCGTGATGGGTGGGATGG 318
QY 61 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 80
DB 319 AACAAITGGGGGTACAAACGGCGTGTGACATTTGGCGAGACCCGATCACTACTAACAA 378
QY 81 ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 100
DB 379 GTATGGCGCGCGTTCGGAATGTACCAACAGCTGGGACCCCAACACGACGAGCGATTCTC 438
QY 101 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPhelysLeuGlyGly 120
DB 439 CGCTCGTCTGTATCAGGGCGCAACGTGGCAAAATAACGCCCTTCCCGTTCAAGTTGGCGGC 498
QY 121 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAspAsnIle 140
DB 499 AACATGCCCGGGCGTGGAAATGGGCGAGCGGCTTGGGTGGATCCAAACATGACAACTT 558
QY 141 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThr 160
DB 559 CTGTATTTCGGCGCGCCGAGCGCAAGGGCTCTGGAGAAGCACAGATTCCGCGCGCGACC 618
QY 161 TrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 180
DB 619 TGGTCCCAGATGACGAACCTTTCCGAGACGTAGGACAGTACATTGCAAAATCCCACTGACAG 678
QY 181 ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSer 200
DB 679 ACCGGCTATCAGAGCGATATTCAAGGCGTCTGTGGTCTGCTTCGACAAAGTCTTCGTCA 738
QY 201 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal 220
DB 739 TCGCTCGGGCAAGCGAGTAAGACCAATTTTGTGGCGGTGGCGGATCCCAATAATCCGCTC 798
QY 221 PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyValAlaProThrGlyPhe 240
DB 799 TTCTGGAGCAGACGCGCGCGAGCTGGGAGCGGTGCCGGGTGCCCGCGCGCGCTTC 858
QY 241 IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn 260
DB 859 ATCCCGCACAGGGCGCTTTTGACCGGTCAACGACGTGCTCTATATTGCCACCAACAAT 918
QY 261 ThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGly 280
DB 919 ACGGGTGGTCCGTATGACGGGAGCTCCGCGACGCTCGAAATCTCGGTGACCTCCGGG 978
QY 281 ThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr 300
DB 979 ACATGGACGCGAATCAGCCGCTACCTTCGACGACGACGCGCAACGCACTACTTTGGTTAC 1038
QY 301 SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer 320
DB 1039 AGCGGCTTCACTATCGACCGCGACACCCCAACAGATTAATGTGCAACCCAGATATCG 1098
QY 321 TrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyValAlaThrTrpThrArgile 340
DB 1099 TGGTGGCGGACACCAATAATCTTCGGAGCACCGGCGGTGGCGAGTGGCGCGATC 1158
QY 341 TrpAspTTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGlu 360
DB 1159 TGGGATGGACGAGTATCCCAATCGAAGCTTGGATATGTGCTTGACATTTCCGGCGAG 1218
QY 361 ProTrpLeuThrPheGlyValGlnProAsnProValProSerProLysLeuGlyTTrp 380
DB 1219 CCTTGGCTGACCTTCGGCGTACAGCGAATCTCCGTAACCGATCGATCGGTCTTACGGA 1278
QY 381 MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGly 400
DB 1279 ATGGATGAACGATGGCAATCGATCCGTTCACTCTGATCGGATGCTCTACGGAACAGGC 1338
QY 401 AlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIle 420
DB 1339 GCGAGCTGTACGCAACAAATGATCTCAGAAAGTGGGACTCCGCGCGCCAGATTCAATATC 1398
QY 421 AlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProSer 440
DB 1399 GCGCGCATGGTCAAAGGATTGGAGGAGCGCGGTAAACGATCTCATCAGCCCGCGCTCT 1458
QY 441 GlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyPheThrHisAlaAspValThr 460
DB 1459 GCGCGCCCGCTCATCAGCGCTCTCGGAGACCTTCGGCGGTTCACCCAGCCGCGGTACT 1518
QY 461 AlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyr 480
DB 1519 GCGGTCCATCGACGATCTTCAGTCAACGGTGTTCAGACCGGCGCACCGCGCTGACTAT 1578
QY 481 AlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSerGlnPro 500
DB 1579 GCGGAATTGAATCCGTCGATCATCGTTCGCTCGGTGGAAGTTTCGATCCATCGAGCAACCG 1638
QY 501 AsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGlu 520
```



Db 1639 AACGACAGCGACGTCGGCTTCTCGACAGACGGCGGCAAGAACTGGTTCCAGGCGAGCGAA 1698  
Qy ProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheVal 540  
Db 1699 CCTGGCGGGTGAAGACGGGGCGGACCGTCCCGCATCGCCGACGGCTCTCGTTTCGTG 1758  
Qy TripAlaProGlyAepProGlyGlnProValValValAlaValGlyPheGlyAsnSerTrp 560  
Db 1759 TGGCTCCCGGGGATCCCGGTGAGCTGTGTGTACGCAGTCGGATTGGCAACTCCTGG 1818  
Qy AlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLys 580  
Db 1819 GCTGCTTCGCAAGGTGTTCCGCCAATGCCAGATCCGCTCAGACCGGGTGAATCCAAAG 1878  
Qy ThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPhe 600  
Db 1879 ACTTTCATGCCCTTATCAATGGAACCTTCTATGAAGCACGGACGGCGGTGACATTC 1938  
Qy GlnProValAlaAlaGlyLeuProSerSerGlyAlaValValMetPheHisAlaVal 620  
Db 1939 CAACGGTCCGGCGGGTCTCCGAGCAGCGGTGCGTGTGTATGTTCCACGGGTG 1998  
Qy ProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsn 640  
Db 1999 CCTGGAAGAAAGCGGATCTGTGGCTCGCTGCATCGAGCGGCTTTACCACTCAACCAAT 2058  
Qy GlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGly 660  
Db 2059 GGCGGACGACGTTGGTCTGCAATCACCGCGGTATCTCCGCGGTGAACGTGGGATTGGT 2118  
Qy LysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGlyGlyVal 680  
Db 2119 AAGTCTGCGCGCGGGTGTGTATACCCAGCGTCTTTGTCTGTCGGCAGATCGAGCGGT 2178  
Qy ThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAspAspGln 700  
Db 2179 ACGGGGGCGTACCGCTCCGACGACTGTGGGACGACCTGGGTACTGATCAATGATGACCCAG 2238  
Qy HisGlnTyrGlyAsnTrpGlyAlaIleThrGlyAspHisAlaAsnLeuArgVal 720  
Db 2239 CACCAATACGGAAATGGGGACAAGCAATCACCGGTGACCAACGGAAATTTACGGCGGGTG 2298  
Qy TyrIleGlyThrAsnGlyArgGlyTyrValTyrGlyAspIleGlyGlyValAlaProSerGly 740  
Db 2299 TACATAGCAGCAAGCGCGCGTGGAAATGTATACGGGGACATTTGGTGGCCCGCTCCGA 2358

## RESULT 4

ABV76941  
ID ABV76941 standard; DNA; 2823 BP.  
XX AC ABV76941;  
XX DT 03-MAR-2003 (first entry)  
XX DE Nucleotide sequence of a xyloglucanase enzyme.  
XX KW Xyloglucanase; enzyme; family 74; glycosyl hydrolase; cellulose fiber;  
KW textile scouring; cellulose fiber processing; ratting; gene; ss.  
XX OS Jonesia sp.  
FH Key Location/Qualifiers  
CDS 1..2823  
FT /\*tag= a  
FT /product= "xyloglucanase enzyme"  
FT sig\_peptide 1..69  
FT /\*tag= b  
FT mat\_peptide 70..2820  
FT /\*tag= c  
XX W0200277242-A2.  
XX PN  
XX PD 03-OCT-2002.  
XX 27-MAR-2002; 2002WO-DK000210.  
XX PF  
XX PR 27-MAR-2001; 2001DK-00000504.  
XX PA (NOVO ) NOVOZYMES AS.  
XX PI Duffner F, Sjöholm C;  
XX DR P-FSDB; ABB99489.  
XX WPI; 2003-092855/08.  
XX P-FSDB; ABB99489.  
XX New xyloglucanase enzyme belonging to family 74 of glycosyl hydrolases,  
XX and endogenous to a bacterium, useful in the textile industry for  
XX improving properties of cellulosic fibers, yarn or fabric.  
XX Claim 11; Page 68-72; 76pp; English.  
XX The present sequence encodes a xyloglucanase enzyme, belonging to family  
XX 74 of glycosyl hydrolases. The enzyme is isolated from Jonesia sp.  
XX DSM14140. The enzyme is useful in processes for machine treatment of  
XX fabrics. It is also useful in the textile industry for improving the  
XX properties of cellulosic fibers, yarn, woven or non-woven fabric, and in  
XX a textile scouring process step. The xyloglucanase enzyme is also useful  
XX in the cellulose fiber processing industry for ratting of fibers such as  
XX hemp, jute, flax and linen. It is useful for preventing binding of  
XX certain soils to the xyloglucan left on the cellulosic material  
XX Sequence 2823 BP; 665 A; 803 C; 764 G; 591 T; 0 U; 0 Other;  
Alignment Scores:  
Pred. No.: 5,01e-124 Length: 2823  
Score: 2063.00 Matches: 386  
Percent Similarity: 67.21% Conservative: 112  
Best Local Similarity: 52.09% Mismatches: 225  
Query Match: 51.11% Indels: 18  
DB: 8 Gaps: 9  
US-09-917-376-3 (1-740) x ABV76941 (1-2823)  
Qy 6 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 25  
Db 130 TATAGCTGTCGACAGCTTCAGATTGTGGCGCGGTTTGTGCTGCGAATGTCCTTAAC 189  
Qy 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAsp 45  
Db 190 CAGAAAGACCCAGGCTTGGTGATGCCGACAGATATTTGGTGGAGCGTACCGGCTCAAT 249  
Qy 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr 65  
Db 250 GACTCCACCGGGCGTGGATCCCTCTTACGGACCATATTTGGTGGATGACTGGAGCCAC 309  
Qy 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85  
Db 310 TCCGGAATTCGTCTCTGGCCACTGACCTGTGTGACCAACCAACCGCGCTGACTTGCAGCA 369  
Qy 86 GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGln 105  
Db 370 GGCAGTACTCCATGGACTGGGATCCCAAAACGGTGTATTTGGTGGTCCGCGGACAAA 429  
Qy 106 GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArg 125  
Db 430 GCGGAGACTGGGAAAGACGATGTCGCTTCCGTGTCGGAGGCAACATGCGGGTTCG 489  
Qy 126 GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 145  
Db 490 GGGATGGGAGAACGCTCGCGGTGACCCCAACAATAATAAAGTCTCTGACTTTGGTGGC 549  
Qy 146 ProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 165  
Db 550 GAATCGGGTAAATGGACTATGGAAGTCACTAGTACGGGAAACCTTGGGGCAAGTACG 609  
Qy 166 AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGly---TyrGln 184



```
Db 610 AGCTTCCCAATGCTGGTAACGTACGTGCGC-----GATGCTTCTGGTGGCGTACACC 660
Qy 185 SerAspIleGlnGlyValValTyrValAlaPheAspLysSerSerSerSerLeuGlyGln 204
Db 661 GCGCAAAACCAAGGTGTTGTGGGTAACTTTTGACCCACATCAGCGAAAGCTGGTCAG 720
Qy 205 AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheThrSerArg 224
Db 721 ACAACACAAACGATCTACGTAGGTGTAGCAGACAAAGACAACAGTTTATCGATCTAC 780
Qy 225 AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHisLys 244
Db 781 GATGCGGGGCAACATGCAACGCTCTCGACACCCACCGGATTCCTGGCGCAAAAG 840
Qy 245 GlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyPro 264
Db 841 GGAGTGTGTGACCATAAAGGTCAGCAACTACATCGCTACCTCAGATACCTGGTGGCCG 900
Qy 265 TyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArg 284
Db 901 TATCAGCGGTGCAAGGGGATGTGGCGCCTAGATATTTCAGTGGCCAGTGGACTCGA 960
Qy 285 IleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThr 304
Db 961 ATCAGCCCAATTCCTCTACCC-----TCATCGAAATTCAGGCTTCGGGTACTCCGGTTAGCC 1017
Qy 305 IleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAsp 324
Db 1018 ATTGACCGCAAAACCCCTGACACCATCATGTTGTTTCCCAAGTCTCTGGTGGCCTGAC 1077
Qy 325 ThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrpAspTrpThr 344
Db 1078 ATGTACGTATATCGTCCACCGACCGGGGAAAACGTGTGCGCTATCTCGGAACGTGAAC 1137
Qy 345 SerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThr 364
Db 1138 GGTTCACAGCCGCTGACCAAGCAGTACCAACCAAGATTACTAGGGGCTCCCTGGCTTGAC 1197
Qy 365 PheGlyValGlnProAsnProValProSerProLysLeuGlyTrpMetAspGluAla 384
Db 1198 TTTGCAATACAGCAAGAGCTGAGCTAAGCTAAACCCCAACTCGGTGGATGACCCAGTCC 1257
Qy 385 MetalIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyr 404
Db 1258 TTCGAGATTGACCCGCATAACTCCGACCGCTTCTTCTATGGCAGCCGCGCGGCTCTAT 1317
Qy 405 AlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetVal 424
Db 1318 GCGGAACGCAACCTCACAACTGGGATAGGGCAAAAGTTGATATCATCTGTCAGGCC 1377
Qy 425 LysGlyLeuGluThrAlaValAsnAspLeuIleSerProProSerGlyAlaProLeu 444
Db 1378 CAAGGGATCGAAGAAACCGCAGCCCAAGACCTTGGCGCTCCCTCGGAACATGGCCTA 1437
Qy 445 IleSerAlaLeuGlyAspLeuGlyPheThrHisAlaAspValThrAlaValProSer 464
Db 1438 TACTCAGCGCTGCGCAGACATCGGTGGCTTCCACCAAGATATCTCTCAGGTGCGCTAAC 1497
Qy 465 Thr----IlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeu 483
Db 1498 AAGTACTACTACAAGAACCCACACCATGACCGGTGACCTCCATTTGCTCCCGCAAGC 1557
Qy 484 AsnProSerIleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArg 503
Db 1558 AAACCGGCAACTGTGTGGTGCAGGCAAA-----TCATTCTGGGGAACCACTCTCA 1611
Qy 504 HisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGly 523
Db 1612 TGGGTGGGTTCACATGATGTGCGGAAACGTGGAAACCTGGCGCCACCGCATCAGGG 1671
Qy 524 ValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro 543
```

```
1672 GTTAAAGCCCGCGCTCCATCACTGTGTGTCAGCTAAATGCTTTCATCCATTGTGTGGGCCCG 1731
Qy 544 GlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSer 563
Db 1732 -----GAAGGTGCGCTCCTCGACGTTCCACCAACAGTGGCTCCTCGTGTCAACGGTG 1785
Qy 564 GlnGlyValProAlaAsnAlaGlnIleAerGserAspArgValAsnProLysThrPheTyr 583
Db 1786 AGCGGCTTCTCACAATGCCAGGTCCGATCAGACCGAGTCAACGCTAATACTCTTTTAC 1845
Qy 584 AlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProVal 603
Db 1846 GGTTCGTGTGACGGAAGTTCTATCACTCCAAATGGTGGGCTTCTCTCACGCGCTCA 1905
Qy 604 Ala---AlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGly 622
Db 1906 GCTTTACGCGGGTTTCCACCGAGTGA-----AACGTTCTGTTCCGTGCGTCCGCGGC 1959
Qy 623 LysGluGlyAspLeuTrpLeuAla-----AlaSerSerGlyLeuTyrHis 637
Db 1960 CGTCAAGGTCACTCTGGCTTGTCTGGAGTATCGGAAGCACCTATGGATGGCGG 2019
Qy 638 SerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnVal 657
Db 2020 TCCACTGATGGCGGGAAGAACTGGACCAAGGTATCTGGAGTTCAGGAAGCGCATCGCGTT 2079
Qy 658 GlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIle 677
Db 2080 GGTTCGTGTAAGCAACCACTAGTAGCGGATACCCAGTGAATTTTCACATCGCCCAAGATT 2139
Qy 678 GlyGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsn 697
Db 2140 GATGAGTTCGAGGAATTTTCGGTCTCGACATGAGGCAAGACCTCGAAACGCATCAAC 2199
Qy 698 AspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeu 717
Db 2200 GACAACCAACACAGTGGTATTGGACTGGCGGCATCAATCAACCGGTGACCCAGATGTCGA 2259
Qy 718 ArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyGlyAla 737
Db 2260 GGGCGCTGTACATTGTGTACCAACGCTGCGGCATATTCTGTTGTGATTCTCTCCACTCCC 2319
Qy 738 Pro 738
Db 2320 CCG 2322
RESULT 5
ADH51583
ID ADH51583 standard; DNA; 2517 BP.
XX AC ADH51583;
XX DT
XX DT
XX DT
XX DT
XX DE T reesei eg16 coding DNA sequence related to ethanol production.
XX EGVI protein; endoglucanase; ethanol production; biomass composition;
XX KW fermentation; cellulohydrolase; detergent production; softening agent;
XX KW cotton fabric; wood pulp degradation; sugar; gene; ds.
XX OS Hypocrea jecorina.
XX Key Location/Qualifiers
FH 1. .2517
FT /tag= a
FT /product= "Mature T reesei EGVI protein"
FT /note= "this sequence contains numerous translation
FT exceptions where the amino acid is given in the protein
FT sequence as J"
XX PN US2003113732-A1.
XX PD 19-JUN-2003.
```



XX 18-DEC-2001; 2001US-00026994.  
 XX PF  
 XX PA  
 XX PR 18-DEC-2001; 2001US-00026994.  
 XX PA (DUNN/) DUNN-COLEMAN N.  
 XX PA (GOED/) GOEDEGEDEBUUR F.  
 XX PA (WARD/) WARD M.  
 XX PA (YAOJ/) YAO J.  
 XX PI  
 XX PI Dunn-Coleman N, Goedegebeur F, Ward M, Yao J;  
 XX WPI; 2004-106460/11.  
 XX DR P-PSDB; ADH51581.  
 XX PT New substantially purified EGVI polypeptide with the biological activity  
 XX PT of endoglucanase, useful for degrading biomass to ethanol.  
 XX PS Example 1; SEQ ID NO 4; 26pp; English.  
 XX CC This invention relates to a novel purified EGVI protein with the  
 XX CC biological activity of an endoglucanase, comprising an amino acid  
 XX CC sequence chosen from a fully defined sequence as given in specification.  
 XX CC The protein is useful for producing ethanol, which involves contacting a  
 XX CC biomass composition with an enzymatic composition containing the protein  
 XX CC of the invention to yield a sugar solution, adding to the sugar solution  
 XX CC a fermentative microorganism, and culturing the fermentative  
 XX CC microorganism under conditions sufficient to produce ethanol, where the  
 XX CC biomass composition may be optionally pretreated. The method further  
 XX CC involves the addition of at least one endoglucanase or cellulohydrolase.  
 XX CC The pretreatment is with a dilute acid. The protein of the invention is  
 XX CC also useful for the production of a detergent composition which may be  
 XX CC used as softening agent and for improving the feel of cotton fabrics and  
 XX CC for degrading wood pulp into sugars. The present sequence is that of the  
 XX CC coding DNA sequence which encodes the T reesei EGVI protein of the  
 XX CC invention.  
 XX SQ Sequence 2517 BP; 509 A; 791 C; 746 G; 471 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 9,01e-93 Length: 2517  
 Score: 1579.00 Matches: 325  
 Percent Similarity: 59.26% Conservative: 123  
 Best Local Similarity: 42.99% Mismatches: 264  
 Query Match: 39.12% Indels: 44  
 DB: 12 Gaps: 20

US-09-917-376-3 (1-740) x ADH51583 (1-2517)

QY 6 TyrThrTrpSerAsnValAlaIle---GlyGlyGlyGlyPheValAspGlyIleValPhe 24  
 DB 61 TTTTCATGGAAGACGTCGAAGCTCGCGCGCGCGCGCTTCGCCCGGCATCATCTTC 120  
 QY 25 AsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrp 44  
 DB 121 CATCCACAGACAAAGGGGTAGCATATGCACGACAGATATTGGCGGCTGTACCGCTC 180  
 QY 45 AspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpVal-----GlyTrp 60  
 DB 181 AAC---GCCGAGACTCATGACCGCCCTCAGCGATGGATTGCTGATAATGTCGGCTGG 237  
 QY 61 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 80  
 DB 238 CACAACTGG-----GGCATCGACCTGTTCGGCTTGTATCCGACGACGATCAAAAG 288  
 QY 81 ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 100  
 DB 289 GTGATGCCGACGTCCGATGATACGAAACAGCTGGATCCGAGTAATGGAGCCATCAIT 348  
 QY 101 ArgSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGly 120  
 DB 349 CGCTCGTCAGACCGCGCGCAACGTGGTCCCTTCACCAACTTGGCCCTTCAAAGTCGGGGT 408

QY 121 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 140  
 DB 409 AACATGCCAGGACGCGGAGCGCGAGAGCGTCTGGCTGTCATCCGCGCAACTCCCAACATC 468  
 QY 141 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThr 160  
 DB 469 ATCTACTTTGGTCTCGCTCAGGAACCGGCTCTCGAAGTCTACGACGCGCGGTGACC 528  
 QY 161 TrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 180  
 DB 529 TTTTCCAAAGTCTCGTCTTACGCCAACTGGGACGTATCATCCAGACCCGAGTGATTC 588  
 QY 181 ThrGlyTyrGlnSerAspIleGlnValValTrpValAlaPheAspLysSerSerSer 200  
 DB 589 AACGCTACAAACAGCAGCACAAGCAAGCACTCATGTGGTTACGTTTCGACTCAACACGACG 648  
 QY 201 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspPro---AsnAsnPro 219  
 DB 649 ACGACCGGGGAGCCACGCTCTCGTATCTTGTGGCAGCGCTGATAACATCATCTGCTTCA 708  
 QY 220 ValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGly 239  
 DB 709 GTCTATGTGACACGATGCCGGCTCCACGTGGAGTGTCTTACCGGGGAGCAGCGGAAA 768  
 QY 240 PheIleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSer 259  
 DB 769 TACTTTCCTCACAAGCGCAAACTGCAGCCAGCAGAGAGGCTTGTATCTGACCTATTTC 828  
 QY 260 AsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSer 279  
 DB 829 GATGCAACAGCGCGCTATGATGGCACACTTGGCTCAGTGTGGAGGTACACATATGCGGG 888  
 QY 280 GlyThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGly 299  
 DB 889 GGAACCTTGGAAAGACATCACCCCTCTCTGTGATCAGATCTA-----TACTTTGGC 939  
 QY 300 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 319  
 DB 940 TTTGGCGGCTTGGCTCGATTTGCAAAAGCCAGGACCCCTGTGTGTTCTTCTTCTTGAAC 999  
 QY 320 SerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArg 339  
 DB 1000 TCTTGGTGGCAGATGCTCAGCTGTTTGGTCCGCTCAGCCGCTCTGGGACCAACATGGAGCCCG 1059  
 QY 340 IleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAla 359  
 DB 1060 ATCTGGCGGTGGCGGAGCTATCCGACTGAGCTATTACTACAGCATCTCAACTCCCAA 1119  
 QY 360 GluProTrpLeuThrPheGly---ValGlnProAsnProProValProSer----- 375  
 DB 1120 GCACCGTGGATCAAGAACAACTTTATCGATGTGACGCGAGTCCCGTCCGATGCTCTC 1179  
 QY 376 ---ProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg 394  
 DB 1180 ATCAAGCGCTCGCTGGATGATGATGCTCTCGAGATTGATCCCAACACGACGACCAACCAC 1239  
 QY 395 MetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSer 414  
 DB 1240 TGGCTCTACGGCAGCGAATGCAATCTTTGGCGGCGACGATCTCAACCACTGGGACAG 1299  
 QY 415 GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAsp 434  
 DB 1300 CGCCACAAATGTGTCATCAATCATCTGCAGACGCGCATCGAGAAATTCCTCGTCCAGGAC 1359  
 QY 435 LeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPhe 454  
 DB 1360 CTGGCTCTGCACCGCGGAGCGAGCTATTGGCGCGAGTTCGAGACGACCAACGCGCTTC 1419  
 QY 455 ThrHisAla-----AspValThrAlaValProSerThrIlePheThrSerProVal 471  
 DB 1420 ACCTTTGCGCAGCAGAAACGACCTCGGACATCGCGCAGACGCTCTGGGCAACGCGCCACA 1479  
 QY 472 PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla 491



```

Db 1480 TGGCCACCTCGACGAGCGTCACTAGCGCGGAACCTGGTCAAGAGCGTGGTCCGGTCC 1539
Qy 492 GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 511
Db 1540 GGCACACCGCGCGCACACAG-----GTGGCCATCTCGTCCGACGGC 1584
Qy 512 GlyAsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyValAla 531
Db 1585 GCGCGAGTGGACATCGACTAGCGCGCCACAGCTCATGAACGCGGCGACGCGTGGCC 1644
Qy 532 AlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValVal 551
Db 1645 TATTCGCGCGACGCGACAGCATCTCTGTGCGACCGCTCGTCCGGC-----GTG 1695
Qy 552 TyrAlaValGlyPheGlyAsnSerTrpAlaAspGlnGlyValProAlaAsnAlaGln 571
Db 1696 CAGCGCTCGAGTTCCAGGCGAGCTTTGCTCTCGAGCTCGCGCGCGCGCGCGTCC 1755
Qy 572 IleArgSerAspArgValAsnProGlyThrPheTyrAlaLeuSerAsnGlyThrPheTyr 591
Db 1756 ATCCCTCGGACAAAGAGACCAACAGCGCTCTCTACGCGCGCTCGGATCGACCTTTTAC 1815
Qy 592 ArgSerThrAspGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGly 611
Db 1816 GTCAGCAAGCACCGCGCAGCAGCTTC-----ACGCGCGG-----CCCAAGCTGGC 1863
Qy 612 AlaValGlyValMet-----PheHisAlaValProGlyLysGluGlyAspLeuTrpLeu 629
Db 1864 AGCGCAGGAGCATCGGGATATCGTCTCACCGGACACCGCGCGGACGTTGTATGTC 1923
Qy 630 AlaAspSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIle--- 648
Db 1924 TCGACCGACGTCGGCATATTCGCTCCACAGCTCGGCGCAGACCTTTGGCCAAAGTCTCC 1983
Qy 649 ThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyr 668
Db 1984 ACCGCTTGACCAACACCTACAGATCCGCTGGGTGTGGGTCTCA-----GGCTGACACG 2040
Qy 669 ProAlaValPheValGlyThrGlyValThrGlyAla-----TyrArgSer 686
Db 2041 ---AACCTGTATGCTTCGGCACC-----GGCCGTCAGGGCTCGCTCTAGCCAGT 2091
Qy 687 AspAspCysGlyThrThrTrpValLeuIleAsnAspGlnHisGlnTyrGlyAsnTrp 706
Db 2092 GGAGACACGCGCGCTCTCGACGACATCCAGGGCTCCAGGGCTTCGGCTCCATCGAC 2151
Qy 707 GlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArgValTyrIleGlyThrAsnGly 726
Db 2152 AGCACCAAGTTCGCGGACGCGGACGCGGACGCGGCGGAGTCTAGTGGGACCAACGCGC 2211
Qy 727 ArgGlyIleValTyr-----GlyAspIleGlyAlaProSerGly 740
Db 2212 CGGGCGCTTTTACGCTCAGGACCTCGCGCGGCGGCGGCGGCGG 2259

```

RESULT 6

```

ADH51580
ID ADH51580 standard; cDNA; 2710 BP.
XX AC ADH51580;
XX
XX
XX 25-MAR-2004 (first entry)
XX Trichoderma reesei eg16 cDNA sequence related to ethanol production.
XX EGVI protein; endoglucanase; ethanol production; biomass composition;
KW fermentation; cellobiohydrolase; detergent production; softening agent;
KW cotton fabric; wood pulp degradation; sugar; gene; ss.
XX Hypocrea jecorina.
XX Location/Qualifiers
FH 98.2557
FT CDS

```

```

/*tag= a
/product= "Mature T reesei EGVI protein"
/Note= "This sequence contains numerous translation
exceptions where the amino acid is given in the protein
sequence as J"
US2003113732-A1.
19-JUN-2003.
18-DEC-2001; 2001US-00026994.
18-DEC-2001; 2001US-00026994.
(DUNN/) DUNN-COLEMAN N.
(GORD/) GOEDEGEBUUR F.
(WARD/) WARD M.
(YAOJ/) YAO J.
Dunn-Coleman N, Goedegebuur F, Ward M, Yao J;
WPI; 2004-106460/11.
P-PSDB; ADH51581.
New substantially purified EGVI polypeptide with the biological activity
of endoglucanase, useful for degrading biomass to ethanol.
Example 1; SEQ ID NO 1; 26pp; English.
This invention relates to a novel purified EGVI protein with the
biological activity of an endoglucanase, comprising an amino acid
sequence chosen from a fully defined sequence as given in specification.
The protein is useful for producing ethanol, which involves contacting a
biomass composition with an enzymatic composition containing the protein
of the invention to yield a sugar solution, adding to the sugar solution
a fermentative microorganism, and culturing the fermentative
microorganism under conditions sufficient to produce ethanol, where the
biomass composition may be optionally pretreated. The method further
involves the addition of at least one endoglucanase or cellobiohydrolase.
The pretreatment is with a dilute acid. The protein of the invention is
also useful for the production of a detergent composition which may be
used as softening agent and for improving the feel of cotton fabrics and
for degrading wood pulp into sugars. The present sequence is that of the
cDNA sequence which encodes the T reesei EGVI protein of the invention.
XX
SQ Sequence 2710 BP; 571 A; 840 C; 782 G; 517 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 9.81e-93 Length: 2710
Score: 1579.00 Matches: 325
Percent Similarity: 59.28% Conservative: 123
Best Local Similarity: 42.99% Mismatches: 264
Query Match: 39.12% Indels: 44
DB: 12 Gaps: 20
US-09-917-376-3 (1-740) x ADH51580 (1-2710)
Qy 6 TyrThrTrpSerAsnValAlaIle---GlyGlyGlyPheValAspGlyIleValPhe 24
Db 101 TTTTCATGGAAGAACGTCGAAGCTCGCGCGCGCGGCTCGTCCCGCGCATCATCTTC 160
Qy 25 AsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArgTrp 44
Db 161 CATCCCAAGACAAAAGCGCTAGCATATGCACGACAGATATTCGCGGCTGTACCGGCTC 220
Qy 45 AspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpVal-----GlyTyr 60
Db 221 AAC---GCCGAGACTCATGGACCGCGCGCGCGGCTCGTCCGATGATGCGGCTGG 277
Qy 61 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 80
Db 278 CACAACTGG-----GGCATCGACGCTGTTCGCGCTTCATCCGAGGACGACGACAAAG 328

```















CC antisense nucleic acid sequence. The nucleic acid sequence of the gene  
 CC encoding the protein is a recombinant sequence having at least one  
 CC mutation as compared to the wild-type gene encoding the protein. The  
 CC transgenic cell or microorganism are useful for producing xanthan gum,  
 CC which are useful for providing formulations and properties, such as long-  
 CC term suspension and emulsion stability in alkaline, acid, and salt  
 CC solutions, temperature resistance, and pseudoplasticity. The present  
 CC sequence encodes an enzyme relating to the present invention.  
 XX  
 SQ Sequence 3668 BP; 570 A; 1181 C; 1292 G; 625 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 4,056-64 Length: 3668  
 Score: 1139.00 Matches: 267  
 Percent Similarity: 50.85% Conservative: 120  
 Best Local Similarity: 35.09% Mismatches: 295  
 Query Match: 28.22% Indels: 79  
 DB: 10 Gaps: 21

US-09-917-376-3 (1-740) x ADD24893 (1-3668)

QY 1 AlaThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAsp 20  
 DB 1217 GCACGTCCGGCCCTACAGTGGCGAGTGTCCGCTTGGCGGTTCCTACC 1276  
 QY 21 GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGly 40  
 DB 1277 GGTGTGCTGTTTCCGCCCGCAACGTGTCTGGCTATGCGCGACCGATGGGTGGC 1336  
 QY 41 MetTyrArgTTPAspAlaAsnGlyArgTTPleProLeuAsnTTPValGlyTTP 60  
 DB 1337 GCGTACCGTGGGATGCGCAGCGCGAGTGGACCGCGCTGACCGACTGGTGGCGCT 1396  
 QY 61 AsnAsnTTPGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 80  
 DB 1397 GAGCACTGGAACCTGATGGGATCGAGCATTCGCGTTCGACCCCGCATCCGATGGG 1456  
 QY 81 ValTTPAlaAlaValGlyMetTyrThrAsnSerTTPAspProAsnAspGlyAlaIleLeu 100  
 DB 1457 CTGTATCTGGCGCGCGACCTATATGCATGAA---CGCGCGGACCTCGCGAGTGTG 1513  
 QY 101 ArgSerSerAspGlnGlyAlaThrTTPGlnIleThrProLeuProPheIleGlyGly 120  
 DB 1514 CGTCTGTTCAACCGCGCGCGACGTTTCGAGCGTTCGCGCTGCGCTTAAAGTGGGTGT 1573  
 QY 121 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAenIle 140  
 DB 1574 AACCACTGGCGCGCGCGCAATGTCGAGCGCTGCGCGTGGACCCCGCATGCGCGCGC 1633  
 QY 141 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTTPArgSerThrAspSerGlyAlaThr 160  
 DB 1634 CTGCTCTGGCTCGCGCGATGCC---GGCTGTGGCGTAGCGACGATCGCGCGCGCAC 1690  
 QY 161 TTPSerGlnMetThrAsnPheProAspValGlyValTTPIleAlaAsnProThrAspThr 180  
 DB 1691 TGGCGCAAGTGGCGTGGTTCGCGACCGCGCGCTGCGCGTGCACCGCGCGCAATCAT 1750  
 QY 181 ThrGlyTyrGlnSerAspIleGlyValValTTPValAlaPheAspLysSerSer 200  
 DB 1751 GTTGGCGCGGACGACGCGGTG---GGGATCGCTTGTGTGTGTCAGCGACCGATGGC 1807  
 QY 201 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal 220  
 DB 1808 AACACGGCTCGCCACACCGCGCATCTAGTGGGCTGTCCACCGACAGACCGCTG 1867  
 QY 221 PheTTPSerArgAspGlyGlyAlaThrTTPGlnAlaValProGlyAlaProThrGlyPhe 240  
 DB 1868 TATGTGTCCAGATGCGCGCGCGTGGTGGCACCGTGGCGCGCAACACCGCGCGCTG 1927  
 QY 241 IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn 260  
 DB 1928 CGCCCCGAGCCACATGCGCGCGCGACGATGGGCAC---TGGTATCTGAGCTATGGCGAC 1984

QY 261 ThrGlyGlyProTyrAspGlySerSerGlyAspValTTPIlePheSerValThrSerGly 280  
 DB 1985 CAGCCCGCGCGCGACCTGATGCCCGGGAGGAGCTTGTGCAAAATTCACGCGCGCACAGGG 2044  
 QY 281 ThrTTPThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr 300  
 DB 2045 CGCTGGCGTGGATCAGCCCGATTCGCGCAG---CCAGCCAGTGGCGGATGGAATTCGCGTGG 2101  
 QY 301 SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer 320  
 DB 2102 GGTGGGTGGCGGTGGATCCGCAACATCCGAGGTGCTGCTGGCCAGCAGCTTCGCCCGT 2161  
 QY 321 TTPTTPProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTTPThrArgIle 340  
 DB 2162 CGCAGCGCGCGCGACGAGCTGATCGCAGCTGGATGGTGGCAAGCACTGGAGCGCGTGG 2221  
 QY 341 TTPAspTTPThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGlu 360  
 DB 2222 -----TTGGCGGATGCGGTGTTTCGATCAGCGCGCGCG 2254  
 QY 361 ProTTPLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTTP 380  
 DB 2255 CCGTGG---ACGCGACATGCCAGCGCGCAC-----TGG 2284  
 QY 381 MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGly 400  
 DB 2285 ATGGGG---GCCTGGCGATCGATCCGTTTCGACGGCAACCATCGCTGTTTCGTGACCGCG 2341  
 QY 401 AlaThrLeuTyrAlaThrAsnAspLeuThrLysTTPAspSerGlyGlyGlnIleHisIle 420  
 DB 2342 TACGGCATCTGGCGCTCGCGCAATCTG-----CAGGATTTCCGCC 2380  
 QY 421 AlaPro-----MetValLysGlyLeuGluThrAlaVal 432  
 DB 2381 GCACCGCAGCGCGCGCTGCGAGTGGTTCACGAGCGGTGGTGAAGAAACCGTCCGC 2440  
 QY 433 AsnAspLeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGly 452  
 DB 2441 CTGGACCTCTCAGCCCGGATGGCTGGCGGCATCTGCTCAGCGCGCTCGCGGATATCGAC 2500  
 QY 453 GlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPhe 472  
 DB 2501 GGCCTTCGGCATGACGACCTGGACCGCGTG---CAGTTGCAGTACGACCGCGCGCGGTG 2557  
 QY 473 ThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGly 492  
 DB 2558 ACCAATGGCGAAAGCATCGATGCGCGCGCGCGCGCGCGCGCGGTGGTGGCGCGCGGT 2617  
 QY 493 SerPhe-AspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyG 512  
 DB 2618 ACCGTGCGCGCGCGCGCAACAGAAATCCGCGCGC---TGTATTACGCGTGTATGGCGG 2674  
 QY 512 YLysAsnTTPPheGln---GlySerGluProGlyGlyValThrThrGlyGlyThrValAl 531  
 DB 2675 CAAGCAATGGATCGCTGCTTTCGAGTGGCGCGCGCGCGCGCGCGCGCGCGCATTCG 2734  
 QY 531 aAlaSerAlaAspGlySerArgPheValTTPAlaProGlyAspProGlyGlnProValVa 551  
 DB 2735 CATTTGGTGGCATGCGCGCGAGGTGGTGGCGCACCGCGCGCGCGCGCAAT-----TG 2788  
 QY 551 lTyrAlaValGlyPheGlyAsnSerTTPAlaAlaSerGlnGlyValProAlaAsnAlaG 571  
 DB 2789 GCGCAGCTCCGATCTTCGGCGCGCAGTGGCAGCGGTGTGAACCCCTCGCGCGCGCGCG 2848  
 QY 571 nIleArgSerAspArgValAsnProLysThrPheTyrAla-----LeuSerAsnGlyTh 589  
 DB 2849 GGTGGTGACCAATCGGTGGACGACGCGCGCTGTATCGGTGATGTCGCGCGCGCGCA 2908  
 QY 589 rPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeu----- 607  
 DB 2909 GCTGTACGAGACGACCGATGCGCGCGCGAGTTTCCGT-----GCGACCGGTGTACAG 2962  
 QY 608 -----ProSerSerGlyAlaVala 613



Db 2363 GGGCAGCCGGCGCGATGAGCGACCCCGCGCGAGCTGCTCCGGACCGCGTGGCGGC 3022  
 Qy 613 lGlyValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaSerSe 633  
 Db 3023 AGGCGTGTGTATCTGGCCAGCCCGGCAAG----- 3053  
 Qy 633 rGlyLeuTrpHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSe 653  
 Db 3054 -GGTGTGATGCGCTGGCAGGAGGT-----GCGCTGAGTGTCTCTCACAGCGGACGA 3106  
 Qy 653 rAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTrpProAlaValPheVa 673  
 Db 3107 AGCAGCTCGTGTGCGCATCGCAAGGCGCTGCGCGCGGTGCGCGCGCTGTATCT 3166  
 Qy 673 lValGlyThrIleGlyValThrGlyValArgSerAspCysGlyThrTr 693  
 Db 3167 GCGCGGCGCGTGCAGGGCGTGGATGGCGTGTTCGCTCCGACGATGCGCGCGCGAGTG 3226  
 Qy 693 pValLeuIleAsnAspGlnHisGlnTrpGlyAsnTrpGlyGlnAlaIleThrGlyAs 713  
 Db 3227 GCAGCGCATCAAGATGAGCGCGACCGCTTCGGGCGC---CCGTACAGCGTGACCGTGA 3283  
 Qy 713 pHisAlaAsnLeuArgArgValTrpIleGlyThrAsnGlyArgGlyIleValTrpGlyAs 733  
 Db 3284 TCCGCGCATTCGCGCGGTGTGTACTTCGCCACCGCGCGCGCATTTCTTACGGCGA 3343  
 Qy 733 p 733  
 Db 3344 T 3344  
 RESULT 9  
 ADH19119  
 ID ADH19119 standard; DNA; 2367 Bp.  
 XX AC ADH19119;  
 XX DT 11-MAR-2004 (first entry)  
 XX DE Geotrichum sp M128 xyloglucan oligosaccharide-degrading enzyme mat DNA.  
 XX KW xyloglucan oligosaccharide degradation; genetic engineering; ds; gene;  
 XX KW mature.  
 XX OS Geotrichum sp. M128.  
 XX FH Key Location/Qualifiers  
 XX FT 1. .2367  
 XX FT CDS /tag= a  
 XX FT /partial  
 XX FT /product= "Geotrichum sp M128 xyloglucan oligosaccharide-  
 XX FT degrading enzyme mature protein"  
 XX FT /note= "No start or stop codon"  
 XX EP1350844-A2.  
 XX PN 08-OCT-2003.  
 XX PD 25-MAR-2003; 2003EP-00251866.  
 XX PF 25-MAR-2002; 2002JP-00083433.  
 XX PR (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.  
 XX PA Yaoi K, Mitsuishi Y;  
 XX PI WPI: 2004-100948/11.  
 XX DR P-PSDB; ADH19120.  
 XX PT Novel xyloglucan oligosaccharide-degrading enzyme with a different  
 XX PT degradation mechanism from known enzymes.  
 XX PS Claim 8; SEQ ID NO 13; 39pp; English.

XX The invention relates to a novel xyloglucan oligosaccharide-degrading  
 CC enzyme with a different degradation mechanism from known enzymes. The  
 CC primary structure and polynucleotide structure of the xyloglucan  
 CC oligosaccharide-degrading enzyme provided by the present invention allows  
 CC high-purity polypeptides having a xyloglucan oligosaccharide-degradation  
 CC activity to be prepared at a low cost through a genetic engineering  
 CC process. The current sequence is that of the Geotrichum sp. M128  
 CC xyloglucan oligosaccharide-degrading enzyme mature DNA of the invention.  
 XX Sequence 2367 BP; 471 A; 797 C; 719 G; 380 T; 0 U; 0 Other;  
 SQ Alignment Scores:  
 Pred. No.: 6.5e-63 Length: 2367  
 Score: 1117.00 Matches: 281  
 Percent Similarity: 50.37% Conservative: 124  
 Best Local Similarity: 34.95% Mismatches: 279  
 Query Match: 27.68% Indels: 120  
 DB: 12 Gaps: 31  
 US-09-917-376-3 (1-740) x ADH19119 (1-2367)  
 Qy 6 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 25  
 Db 10 TACGAGTTTCAAGAATGTCGCGATCGCGCGCGGTATACATTACCGGATTCGCGCAC 69  
 Qy 26 GluGlyAlaProGlyIleLeuTrpValArgThrAspIleGlyGlyMetTrpArgTrpAsp 45  
 Db 70 CCAAGACCAAGACCTGCTGTACCGCGCGACGACATTTGGCGCGGTACCGCTGGGAC 129  
 Qy 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTrp 65  
 Db 130 GCAGCGACGTCCAAGTGGATCCGCTCAACGACTTTATCGAGCGCGACGACATGAACATT 189  
 Qy 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaVal 85  
 Db 190 ATGGGCACCGAGTCGATCGCGTGCAGCCCAACACCCCGCAGAGCTGTACCTCGCGCAG 249  
 Qy 86 GlyMetTrpThr---AsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAsp 104  
 Db 250 GGGCGCTATGCGGCGCAGAGTGG-----GCGGCGTTCTTATGTGTCGGAAGAC 297  
 Qy 105 GlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGly 124  
 Db 298 CGCGCCAGTCGTTTACATCTACAGTCGCGGTCCCGATCGCGCGCCACACGACATGGGA 357  
 Qy 125 ArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTrpPheGly 144  
 Db 358 CGCAACAATGCGAGCGCTCGCTGTCAACCGCTTCAACTCGAAGAGGTCTGGATGGGT 417  
 Qy 145 AlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMet 164  
 Db 418 ACGCGTACA---GAGGGTATCTGGAAGAGTTGGGACCGCGCCCAAGACCTGGCAACACGTC 474  
 Qy 165 ThrAsnPheProAspValGlyThrTrpIleAlaAsnProThrAspThrThrGlyTrpGln 184  
 Db 475 ACGTCCATCCCGACCGCTTC-----ACCAACGGTATCGGATACACG 516  
 Qy 185 SerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGln 204  
 Db 517 TCG-----GTCATTTTCGACCCG-----GAA 537  
 Qy 205 AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArg 224  
 Db 538 CGTAATGGCACCATCTACCGAGCGCGACTGCGCGCGCGAGGCG---ATGTACGTTCAGCAC 594  
 Qy 225 AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIle----- 241  
 Db 595 GACGCGGTGTCTCGTGGGAGCAGTGGCGGCGCGCGCTCCAGCTGCTCAACAGGAC 554  
 Qy 242 -----ProHisLysGlyVal 246  
 Db 655 ACGGGCGGTTCGCCGACAGAGCCCGGTGATCGCGCGCGCGCATGAAGTCGCT 714







DR WPI; 2004-100948/11.  
 XX P-PSDB; ADH19118.  
 PT Novel xyloglucan oligosaccharide-degrading enzyme with a different  
 XX degradation mechanism from known enzymes.  
 PS Claim 11; SEQ ID NO 11; 39pp; English.  
 XX The invention relates to a novel xyloglucan oligosaccharide-degrading  
 CC enzyme with a different degradation mechanism from known enzymes. The  
 CC primary structure and polynucleotide structure of the xyloglucan  
 CC oligosaccharide-degrading enzyme provided by the present invention allows  
 CC high-purity polypeptides having a xyloglucan oligosaccharide-degradation  
 CC activity to be prepared at a low cost through a genetic engineering  
 CC process. The current sequence is that of the *Geotrichum* sp. M128  
 CC xyloglucan oligosaccharide-degrading enzyme precursor cDNA of the  
 CC invention.  
 XX  
 SQ Sequence 2646 BP; 550 A; 866 C; 787 G; 443 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 7.38e-63 Length: 2646  
 Score: 1117.00 Matches: 281  
 Percent Similarity: 50.37% Conservative: 124  
 Best Local Similarity: 34.95% Mismatches: 279  
 Query Match: 27.68% Indels: 120  
 DB: 12 Gaps: 31  
 US-09-917-376-3 (1-740) x ADH19117 (1-2646)  
 QY 6 TyrThrTrpSerAsnValAlaIleGlyGlyPheValAspGlyIleValPheAsn 25  
 DB 198 TACGAGTCAAGATGCGCATCGGCGCGCGGTACATTACCGGATGTGCGGCAC 257  
 QY 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArgTrpAsp 45  
 DB 258 CCAAGACCAAGGACCTGCTACGCGCGCGCGACATATGGCGCGCTACCGCTGGGAC 317  
 QY 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTyrAsnAsnTrpGlyTyr 65  
 DB 318 GCAGGACAGTCCAAAGTGGATCCGCTCAACACATTATCGAGGCGCGACGATGAACATT 377  
 QY 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85  
 DB 378 ATGGGACCGAGTGCATGCGCTGGACCCCAACACCCCGACAGCTGTACTCGCGCAG 437  
 QY 86 GlyMetTyrThr---AsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAsp 104  
 DB 438 GGGCGCTATGTCGCGACAGTGG-----GGCGGTCTCTATGTGTCGCAAGAC 485  
 QY 105 GlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGly 124  
 DB 486 CGCGGCCAGTCGTTTACAACTACGAGTCGCGCGTTCGCGATGGGCGCCCAACGACATGGGA 545  
 QY 125 ArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGly 144  
 DB 546 CGCAACATGCGAGCGGCTCGTGTCAACCGCTTCAACTCGAACGAGGTCTGGATGGGT 605  
 QY 145 AlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaIleThrTrpSerGlnMet 164  
 DB 606 ACGGTACA---GAGGGTATCTGGAAGAGTTCGGACCGCGCCCAAGACCTGGACAAACGTC 662  
 QY 165 ThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrGlyTyrGln 184  
 DB 663 ACGTCCATCCCGGACGGTTC-----ACCAACGGTATCGGATACACG 704  
 QY 185 SerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGln 204  
 DB 705 TCG-----GTCATTTTCGACCC-----GAA 725  
 QY 205 AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArg 224  
 DB 726 CGTAAATGGCACCATCTACGCGGCGGCGACTGCCCGCGCAGGGC---ATGTAGTCAACGCAC 782

QY 225 AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIle----- 241  
 DB 783 GACGCGGTGTCTCGTGGAGCCAGTGGCGGCGCAGCGTCCAGCTGGTGGTCAACAGGACC 842  
 QY 242 -----ProHisLysGlyVal 246  
 DB 843 ACGGCGCGGTTCCTCGGACAAGAAGCCCGCTCGATCGCGCCGACCCATGAAGTCGT 902  
 QY 247 PheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAsp 266  
 DB 903 CTCACCCCC-----AACTTCTCTACGTGACTTAGCGGACTACCTGGTCAATGGGC 956  
 QY 267 GlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArgile--- 285  
 DB 957 GTCAGTTCGGGAGTCTGGCGCCAGAACCGCACCTCGGCGCGCTGGGACGACATTACT 1016  
 QY 286 -----SerProValProSerThrAspThrAlaAsnAspTyrPheGly 299  
 DB 1017 CCCGCGTCGGCAACTCTCGTCCCTGCCCTACAAACACACAGACGTTCCCTCGGCGGGA 1076  
 QY 300 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 319  
 DB 1077 TTTTGGGTCTCAGCGTCGACGCGACCAACCCCAACCGTCTCGTGTCTATCACC---CTC 1133  
 QY 320 SerTrpTrpProAspThrIle-----IlePheArgSerThrAspGlyGlyAlaThr 336  
 DB 1134 GACCGGACCCCGGACCGCCCTCGACGACATCTACTCTCAACCGATCGCGGCGGAC 1193  
 QY 337 TrpThrArgIleTrpAspTrpThrSer-----TyrPro 347  
 DB 1194 TGGAGGACGTCACCCAGCTCTCGTCCCGTCAACCTCGAAGTAACCTGGGCGCACCCG 1253  
 QY 348 AsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpIleThrPheGlyVal 367  
 DB 1254 ACTAACCGCGCGCGGTAC---AAGACGCGACGCTGTTCGTTGGTCTCGACTTCAACAAC 1310  
 QY 368 GlnPro-----AsnProProValProSerPro-----LysLeuGly 379  
 DB 1311 GGTCCCAAGTGGGGGGATACGGTCGCGCGACGGTACGCCCGGCTCACCAAGTTTGGC 1370  
 QY 380 TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr 399  
 DB 1371 TGGTGGATGAGCGCTGTCTTATCGATCGTTTCAACCCCGACGACCTGTATGTACGCGACG 1430  
 QY 400 GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLys-----TrpAspSerGly 415  
 DB 1431 GGGGCGACCATCTGGGCGACCGACACGCTCTCCCGTGTGAGAGGAGACTGG----- 1481  
 QY 416 GlyGlnIleHisIleAlaPro-----MetValLysGlyLeuGluGluThrAla 431  
 DB 1482 -----GGCCGAGCTGGTACCTCCAGATCGAGTATCGAGGAGATGCG 1526  
 QY 432 ValAsnAspLeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeu 451  
 DB 1527 ATCTGTGCTCGCTCGCCCAAGAGCGCGCGCTCTCTGTGGGCATCGGTGACATT 1586  
 QY 452 GlyGlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProVal 471  
 DB 1587 AGCGCATGAAGCACGACGACCTCCCAAG---CCCCAGAAGATGTTTGTGCGGCCCGAC 1643  
 QY 472 PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla 491  
 DB 1644 TTCTCCAACTCGACGACGATCGACCTCGGGGCAACTTCCCAACAGTGTGTGTCGCGGCC 1703  
 QY 492 GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 511  
 DB 1704 GGATCTCGGACACGATACGACGCGGTGCGCGGTGCGTACGCGACTGACGGC 1763  
 QY 512 GlyLysAsnTrp-----PheGlnGlySerGluProGlyGlyValThrThr----- 526  
 DB 1764 GGAGACGCGTGGACCATCTTCCCTACTTCCCTCTCTGCGCATGAACCGGACCACTACCAG 1823



```
QY 527 GlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro-----Gly 544
DB 1824 GGCACGAGTTCAGTTCGACGCGGCGGACGATCGTGTGTCGACCAAGCTTGAC 1883
QY 545 AspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGln 564
DB 1884 GAGCAGGCTCGGACCGTGTACTCGCAGCAGTATGCAAGACGTGTCT-----1934
QY 565 GlyValProAla-----AsnAlaGlnIleArgSerAspArgValAsn 578
DB 1935 ---GTTCCGCTGGCGACTGAAGGCCGAGACTGCAATGTGCTCTCGCAAGTCCAG 1991
QY 579 ProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyVal 598
DB 1992 GATGCGACGTCTACGCTACCGATGGCGGCAAGTTCCTGCTCGACCGCGCGGAG 2051
QY 599 ThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHis 618
DB 2052 TCGTATGCGCCCAAGGCGCGGACTTGTCACT-----GGCACATCGCTCATGCTGCC 2105
QY 619 AlaValProGlyLysGluGlyAspLeuTrpLeu-----AlaAlaSerSerGlyLeuTyrHis 637
DB 2106 GTGAACCCCTGGGTGGCGGACGCTCTGGGTGCTGTCCGAGGGGCTCTCTCCAC 2165
QY 638 SerThrAsnGlyLysSerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsn 656
DB 2166 TCGACCGACTTGGCGCTCGTTCACGAGGTAGTACCGCAACGCGACCTCGTGAGC 2225
QY 657 ValGlyPheGlyLysSer-----AlaProGlySerSerTyrProAlaValPhe 672
DB 2226 GTGCGGCCCCCAAGTCCAGTCGCGCAGAGAGGTAGCGCGCCCTCCGCGTCTTC 2285
QY 673 ValValGlyThr-----IleGlyValThrGlyAlaTyrArgSerAspCysGly 690
DB 2286 ATCTGGGACCGACAGCCTCGAGCGACATCGGCTGTACCGTCCGACCAACGCGC 2345
QY 691 ThrThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIle 710
DB 2346 AGCACCTGGACGCGGTCAATGACGAGGACCAACTACTCGGC-----CCCAACCATGATC 2402
QY 711 ThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyValGlyIleVal 730
DB 2403 GAGGCGGACCCCAAGGTTACGGGCGGTGTATCTAGGCACCAACGCGCGGTATCGTG 2462
QY 731 TyrGlyAspIle 734
DB 2463 TAGCGGACCTT 2474

RESULT 11
ID ADH19123
XX ADH19123 standard; DNA; 2481 BP.
AC ADH19123;
XX
DT 11-MAR-2004 (first entry)
DE Geotrichum sp M128 xyloglucan oligosacch-degrading enzyme-derived DNA.
XX xyloglucan oligosaccharide degradation; genetic engineering; ds; gene;
KW His tag.
OS Synthetic.
XX Geotrichum sp. M128.
XX
FH Key Location/Qualifiers
FT 1..2481
FT /tag= a
FT /product= "Geotrichum sp M128 xyloglucan oligosaccharide-
XX degrading enzyme-derived protein"
FN EP1350844-A2.
XX
PD 08-OCT-2003.
```

```
XX 25-MAR-2003; 2003EP-00251866.
PF
XX 25-MAR-2002; 2002JP-00083433.
PR
XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA
XX Yaoi K, Mitsuishi Y;
PI
XX WPI; 2004-100948/11.
DR P-PSDB; ADH19124.
XX
PT Novel xyloglucan oligosaccharide-degrading enzyme with a different
PT degradation mechanism from known enzymes.
XX
PS Example 7; SEQ ID NO 17; 39pp; English.
XX
CC The invention relates to a novel xyloglucan oligosaccharide-degrading
CC enzyme with a different degradation mechanism from known enzymes. The
CC primary structure and polynucleotide structure of the xyloglucan
CC oligosaccharide-degrading enzyme provided by the present invention allows
CC high-purity polypeptides having a xyloglucan oligosaccharide-degradation
CC activity to be prepared at a low cost through a genetic engineering
CC process. The current sequence is that of the Geotrichum sp. M128
CC xyloglucan oligosaccharide-degrading enzyme-derived DNA of the invention
CC which comprises the wild-type enzyme sequence plus a C-terminal region
CC including a Histidine tag originating from vector pET29a(+).
XX
SQ Sequence 2481 BP; 496 A; 836 C; 748 G; 401 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1,07e-62 Length: 2481
Score: 1114.00 Matches: 281
Percent Similarity: 50.25% Conservative: 123
Best Local Similarity: 34.95% Mismatches: 280
Query Match: 27.60% Indels: 120
DB: 12 Gaps: 31

US-09-917-376-3 (1-740) x ADH19123 (1-2481)

QY 6 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 25
DB 13 TAGCAGTTCAAGAATGTCGCGATCGCGCGGGGTACATTACCGGGATTGTCGCGAC 72
QY 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAsp 45
DB 73 CCAAGACCAAGGACCTGTGTACGCGCGCACGACATTGCGCGCGGTACCGCTGGAC 132
QY 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnTrpGlyTyr 65
DB 133 GCAGGACGTCCTCAAGTGGATCCCGCTCAACGACTTATTCAGGCGCGGACATGAACATT 192
QY 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85
DB 193 ATCGGACCGCAGTCTGATCGCGCTGGACCCCAACACCCGACAGGTGTACTCGCGAG 252
QY 86 GlyMetTyrThr---AsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAsp 104
DB 253 GGGCGCTATGTCGCGCAGCAGTGG-----GCGGCGTTCCTATGTGTCGGAAGAC 300
QY 105 GlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGly 124
DB 301 CGCGGCGAGTCGTTTCAATCTACGAGTCGCCGTTCCTCGATGGCGGCCAACGACATGGGA 360
QY 125 ArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGly 144
DB 361 CGCAACAATGGCGAGCGCTCGCTGTCAACCCGTTCACTCGAACGAGGTCTGGATGGGT 420
QY 145 AlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMet 164
DB 421 ACGCGTACA---GAGGGTATCTCGAAGAGTTCGACCGCGCCCAAGACCTGGACAACGTC 477
QY 165 ThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrGlyTyrGln 184
```



```
Db 478 ACCTCCATCCGGACGGCTTC-----ACCAACGGTATCGGATACACG 519
Qy 185 SerAspIleGlnGlyValValTrrpValAlaPheAspLysSerSerSerLeuGlyGln 204
Db 520 TCG-----GTCAATTTTCGACCCC-----GAA 540
Qy 205 AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrrpSerArg 224
Db 541 CGTAATGGCACCATCTACGGAGCGCGACTGCCCCGAGGGC---ATGTACGTCACCGAC 597
Qy 225 AspGlyGlyAlaThrTrrpGlnAlaValProGlyAlaProGlyPheIle----- 241
Db 598 GACGGCGGTCTCGTGGAGCAGTGGCGGCCCGCTCCAGCTGGCTCAACAGGACC 657
Qy 242 -----ProHisLysGlyVal 246
Db 658 ACGGCGCGTTCGCGGACAAGACCGCGCTCGATCGCGCGCGACGCCATGAAGTCGCT 717
Qy 247 PheAspProValAsnHisValLeuTrrpIleAlaThrSerAsnThrGlyGlyPrrpTyrAsp 266
Db 718 CTCACCCCC-----AACTTCCTCTACGTGACTTACCGCCGACTACCTCGTTCATGGGC 771
Qy 267 GlySerSerGlyAspValTrrpLysPheSerValThrSerGlyThrTrrpThrArgIle--- 285
Db 772 GTACGTTCCGCAAGTCTGGCGCAGAACCGCACCTCGGGCGCTGGGACGACATTACT 831
Qy 286 -----SerProValProSerThrAspThrAlaAsnAspTrrpPheGly 299
Db 832 CCCCGCGTCGGCAACTCGTCGCTGCGCCCGGTACAAACACGACAGCTTCCCTGCGGGCGGA 891
Qy 300 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 319
Db 892 TTTTCGGGTCTACGCTCGACGCGACCAACCCCAACCGTCTCGTCGTCATCAC---CTC 948
Qy 320 SerTrrpTrrpProAspThrIle-----IlePheArgSerThrAspGlyGlyAlaThr 336
Db 949 GACCGCACCCGACCGCGCTCGACGAGCATCTACCTCTCAACCGATGCGCGGCGGACC 1008
Qy 337 TrrpThrArgIleTrrpAspTrrpThrSer-----TyrPro 347
Db 1009 TGGAAGGAGGTACCCAGCTCTCGTCCCGTCCCAACCTCGAAGGTAACTGGGGCCACCG 1068
Qy 348 AsnArgSerLeuArgTrrpValLeuAspIleSerAlaGluProTrrpLeuThrPheGlyVal 367
Db 1069 ACTAACCGCGCGCGGTAC---AAGGACGGCACGCGCTGTTCGTGGCTCGACTTCAACAAC 1125
Qy 368 GlnPro-----AsnProProValProSerPro-----LysLeuGly 379
Db 1126 GGTCCCGAGTGGGGGATACGTCGCGCGCACCGTACGCCCGCGCTCACCAAGTTGGC 1185
Qy 380 TrrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTrrpGlyThr 399
Db 1186 TGGTGGATGACGCGTGTATCGATTATCGATCCGTTCAACCCCGAGCACCTGATGTACGGCAG 1245
Qy 400 GlyAlaThrLeuTrrpAlaThrAsnAspLeuThrLys-----TrrpAspSerGly 415
Db 1246 GGGCGCACCATCTGGCGCGACCGACGCTCTCCGCTGTTCGAGAAGGACTGG----- 1296
Qy 416 GlyGlnIleHisIleAlaPro-----MetValLysGlyLeuGluGluThrAla 431
Db 1297 -----GGCCCGAGCTGGTACTCTCCAGATCGACGGTATCGAGAGAAATGCG 1341
Qy 432 ValAsnAspLeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeu 451
Db 1342 ATCTGTGCTCGCTCGCCCAAGAGCGCGCGGCTCTCTGTCGGGCATCGGTGACATT 1401
Qy 452 GlyGlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProVal 471
Db 1402 AGCGGCATGAAGCACGACGACCTCACCAAG---CCCCAAGAAGATGTTTGGTGGCGGCCAG 1458
Qy 472 PheThrThrGlyThrSerValAspTrrpAlaGluLeuAsnProSerIleIleValArgAla 491
```

```
Db 1459 TTCTCCAACTCGACGATCGACGCTGCGGGCAACTTCCCAACGTTGTGTCTCGTCGCGCC 1518
Qy 492 GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 511
Db 1519 GGATCCTCGGACACGAGTAGCAGCGCGCTGCGCGCGGTGCGTACGCGACTGACGCGC 1578
Qy 512 GlyLysAsnTrrp-----PheGlnGlySerGluProGlyGlyValThrThr 526
Db 1579 GGAGACGGTGGACCATCTTCCCTACCTGCGCTCTTGGCATGAACGCGAGCCACTACCAG 1638
Qy 527 GlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrrpAlaPro-----Gly 544
Db 1639 GGCAGCAGCATGTCAGTCGACGCGAGCGGACGATCGTGTGTCGCCAACCAAGCTTGAC 1698
Qy 545 AspProGlyGlnProValValTrrpAlaValGlyPheGlyAsnSerTrrpAlaAlaSerGln 564
Db 1699 GAGCAGCGCTCGGACCGTGTACTCTGCACGACTATGCAAGACGCTGTCT----- 1749
Qy 565 GlyValProAla-----AsnAlaGlnIleArgSerAspArgValAsn 578
Db 1750 ---GTTCCCGCTGGGACCTGAAGGCCGAGACTGCCAATGTCTCTCGGACAAAGTCCAG 1806
Qy 579 ProLysThrPheTrrpAlaLeuSerAsnGlyThrPheTrrpArgSerThrAspGlyVal 598
Db 1807 GATGCCACGTTCTACGCTACCGATCGGCGGCAAGTCTTCTCGTCTCGACGCGCGGGAAG 1866
Qy 599 ThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHis 618
Db 1867 TCGTATGCGGCCAAGGCGCGGACTTGTCACT-----GGCACATCGCTCATGCTGCC 1920
Qy 619 AlaValProGlyLysGlyAspLeuTrrpLeu---AlaAlaSerSerGlyLeuTrrpHis 637
Db 1921 GTGAACCCCTGGGTGGCGCGGACGCTGGGTGCTGTCCCGAGGGCGGTCTCTTCCAC 1980
Qy 638 SerThrAsnGlySerSerTrrpSerAlaIle---ThrGlyValSerSerAlaValAsn 656
Db 1981 TCGACCGACTTGGCGCTCGTTCACGAGGTAGTACTCCGCCAACGCGACCCCTCGTGAGC 2040
Qy 657 ValGlyPheGlyLysSer-----AlaProGlySerSerTrrpProAlaValPhe 672
Db 2041 GTCGCGCGCCCAAGTCCAGTCGAGCGGCAAGAGGTAGCGCGCCCTCCGCGGTCTTC 2100
Qy 673 ValValGlyThr-----IleGlyGlyValThrGlyAlaTrrpArgSerAspAspCysGly 690
Db 2101 ATCTGGGACCCGACGACCTGGAGCGACATCGCCTGTACCGTCCGACGACCAACGCG 2160
Qy 691 ThrTrrpValLeuIleAsnAspGlnHisGlnTrrpGlyAsnTrrpGlyGlnAlaIle 710
Db 2161 AGCACCTGGACGCGCTCAATGACGAGGACCAACTACTCGGGC---CCCACCATGATC 2217
Qy 711 ThrGlyAspHisAlaAsnLeuArgValTrrpIleGlyThrAsnGlyArgGlyIleVal 730
Db 2218 GAGGCGGACCCCAAGGTCTACGGCGCGGTATCTTAGGCACGACGCGCGGTATCGTG 2277
Qy 731 TyrGlyAspIle 734
Db 2278 TACGCCGACCTT 2289
RESULT 12
ADR90297
ID ADR90297 standard; DNA; 2268 BP.
XX ADR90297;
AC ADR90297;
DT 16-DEC-2004 (first entry)
XX Geotrichum sp. M128 xyloglucanendohydrolase mature DNA.
XX endo-type xyloglucanendohydrolase;
KW xyloglucan-specific endo-b-1,4-glucanase; glucose; xylose; xyloglucan;
KW plant cell differentiation; mature; ds; gene.
XX Geotrichum sp. M128.
OS.
```







```
QY 443 ProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaVal 462
Db 1351 CACCTCTTCCGGGATGACGACCTCGCGGCATCGCCATGATGACTTC---TCCGTC 1407
QY 463 ProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGlu 482
Db 1408 CCGCAGCCCATGTACAGCAAGCCGACCTTCAGCAGCAGCAGCGCGCTCGACTTTCCGGGC 1467
QY 483 LeuAsnProSerIleAlaValArgAlaGlySerPheAsp-----ProSerSerGlnPro 500
Db 1468 CGCGCGGCCAATGTCTCGCGCGCTCGCGCGCAACGACCATCTCTGACGCGCGCTCGCG 1527
QY 501 AsnAspArgHisAlaPheSerThrAspGlyGlyLysAsnTrp-----PheGlnGly 518
Db 1528 GGGTCACGCGCGCGCTACAGCAACCAACAGCGCGCAGCTGGACGCTCTTTCAGACC 1587
QY 519 SerGluProGly---GlyValThrThrGlyThrValAlaAlaSerAlaAspGlySer 537
Db 1588 TGTGTCCCGCAGCTCGGTAAACGCGCGCAGCATCGCGGTGCGCGCGCAGCGCAAG 1647
QY 538 ArgPheValTrpAlaProGlyAsp---ProGlyGlnProValValThrAlaValGlyPhe 556
Db 1648 ACGTTTGTGTGTGTCGCCAGCAGCGCAGCGCAGCGCGCTACACATCGAGCGCACTAC 1707
QY 557 GlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGln---IleArgSerAsp 575
Db 1708 GGCAGACGTGGACTCGCGCGAGCGCTCTGACAGCAGCAGCAGCAGTATCGCCCGCAC 1767
QY 576 ArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAsp 595
Db 1768 CGCGTCAGCGCCACACATCTTACGTAGCTGCGAGGCGGACTTCTTCTCGACCGCAT 1827
QY 596 GlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyVal 615
Db 1828 GCGCGCAAGAGCTACACCAAGAGCGCAACCGCGCTCCGCTCGTGGACG----- 1878
QY 616 MetPheHisAlaValProGly-----LysGluGlyAspLeuTrpLeuAlaAla 631
Db 1879 ---TACACGGGACCGCGCTGACGAGCAACTGCGTGGCGGAGCTGTGGTCTCCGTC 1935
QY 632 SerSer---GlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGly 650
Db 1936 AAGGCGGTGGCATCTACACAGCAGCAGCAGCTTTGGCAACACATTCACCGCGCTCGCGGC 1995
QY 651 ValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAla 670
Db 1996 TCGGCTCCAGTCTCAAT-----CCCGCT 2019
QY 671 ValPheValValGlyThr-----IleGlyGly 679
Db 2020 GTGTTTACATCGCGCGCGCGCAGCGCCCAACGCCACCGAGACGCTCTTCTCTCGGGGC 2079
QY 680 Val-----ThrGlyAlaTyrArgSerAspCysGlyThrThrTrp 693
Db 2080 ATCCGCTCGGCGTGCAGCGCGCGCGCTGTATATGTCGACCGCAACGCGGGTGTGG 2139
QY 694 ValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAsp 713
Db 2140 ACGCGCCTCAACAGCAGCGCGCACAACATATGTTGGC---GCCACCGTGTATCAGCGCGCAT 2196
QY 714 HisAlaAsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleVal 730
Db 2197 CCCCAGATCTACGCGCGGCTGTACATTGGCATGAAACGCGCGCGGCGCATCATC 2247
RESULT 13
ID ADR90301
XX ADR90301 standard; DNA; 2274 BP.
AC ADR90301;
XX
XX
DT 16-DEC-2004 (first entry)
XX
DE Geotrichum sp. M128 xyloglucanendohydrolase mature DNA + Met.
```

```
XX endo-type xyloglucanendohydrolase;
KW xyloglucan-specific endo-b-1,4-glucanase; glucose; xylose; xyloglucan;
KW plant cell differentiation; mature; ds; gene.
OS Geotrichum sp. M128.
XX Synthetic.
FH Key Location/Qualifiers
CDS 1..2274
FT /*tag= a
FT /product= "Geotrichum sp. M128 xyloglucanendohydrolase
FT mature protein + Met"
PN JP2004261037-A.
XX
PD 24-SEP-2004.
XX
XX 28-FEB-2003; 2003JP-00053286.
XX
XX 28-FEB-2003; 2003JP-00053286.
XX
XX (DOKU-) DOKURITSU GYOSEI HOJIN SANGYO GIJUTSU SO.
XX
XX WPI; 2004-665466/65.
XX P-PSDB; ADR90302.
XX
XX Novel xyloglucan hydrolase derived from Geotrichum genus, having ability
XX to hydrolyze beta-1, 4-glycoside linkage of glucose residue in
XX xyloglucan, useful for elucidating structure or function of xyloglucan.
XX
XX Example 6; SEQ ID NO 18; 86pp; Japanese.
XX
XX The invention relates to a novel xyloglucanendohydrolase (xyloglucan-
XX specific endo-b-1,4-glucanase) enzyme derived from the Geotrichum genus
XX and having the ability to hydrolyse the beta-1,4-glycosidic linkage of a
XX glucose residue, but not that of a xylose residue contained in a
XX xyloglucan. The polypeptide of the invention may be useful for
XX elucidating the structure or function of a xyloglucan having an important
XX role in the differentiation of a plant cell. The current sequence is that
XX of the Geotrichum sp. M128 xyloglucanendohydrolase mature DNA of the
XX invention which has DNA encoding an additional Met residue at the N-
XX terminal.
SQ Sequence 2274 BP; 415 A; 795 C; 719 G; 345 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 5,25e-54 Length: 2274
Score: 979.00 Matches: 264
Percent Similarity: 48.43% Conservative: 122
Best Local Similarity: 33.12% Mismatches: 279
Query Match: 24.26% Indels: 132
DB: 13 Gaps: 33
US-09-917-376-3 (1-740) x ADR90301 (1-2274)
QY 11 ValAlaIleGlyGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGly 30
Db 25 GTGACGATCTCGGGTGGCGGCTTCTCGGGGCTCGTGGCACACCCCGCAGGAGGAC 84
QY 31 IleLeuTyrValArgThrAspIleGlyMetTyrArgTrpAspAlaAlaAsnGlyArg 50
Db 85 CTGATCTACGACGACGACGACATTTGGCGCACGTACAGTGGACCGCGCCCAAGTGGAG 144
QY 51 TrpIleProLeuLeuAspTrpVal-----GlyTyrAsnAsnTrpGlyTyr 65
Db 145 TGGGAGCCCATCAGCGACTTTATCATCAACACGCGCTCGCGGCGACCGCGCCACCTG 204
QY 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnIlyValTrpAlaAlaVal 85
Db 205 CTCGCGCAGGAGCATTTGCGCTCGACCCCGCACCGCGCGCTGTACCTCGCGCAA 264
QY 86 GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGln 105
```







```

XX ADR90295;
AC
XX
XX 16-DEC-2004 (first entry)
DT
XX
XX Geotrichum sp. M128 xyloglucanendohydrolase precursor DNA.
DE
XX
XX endo-type xyloglucanendohydrolase;
KW xyloglucan-specific endo-b-1,4-glucanase; glucose; xylose; xyloglucan;
KW plant cell differentiation; precursor; ds; gene.
XX
XX Geotrichum sp. M128.
OS
XX
XX Key Location/Qualifiers
FH CDS 202..2532
FT /*tag= b
FT /product= "Geotrichum sp. M128 xyloglucanendohydrolase
FT precursor protein"
FT 202..261
FT /*tag= a
FT /label= Signal_peptide
FT mat_peptide 262..2529
FT /*tag= c
FT /product= "Geotrichum sp. M128 xyloglucanendohydrolase
FT mature protein"
XX
XX JP2004261037-A.
PN
XX
XX 24-SEP-2004.
PD
XX
XX 28-FEB-2003; 2003JP-00053286.
PF
XX
XX 28-FEB-2003; 2003JP-00053286.
PR
XX
XX (DOKU-) DOKURITSU GYOSBI HOJIN SANGYO GIJUTSU SO.
PA
XX
XX WPI; 2004-665466/65.
DR P-PSDB; ADR90296.
DR
XX
XX Novel xyloglucan hydrolase derived from Geotrichum genus, having ability
PT to hydrolyze beta-1, 4-glycoside linkage of glucose residue in
PT xyloglucan, useful for elucidating structure or function of xyloglucan.
XX
XX Claim 10; SEQ ID NO 12; 86pp; Japanese.
PS
XX
XX The invention relates to a novel xyloglucanendohydrolase (xyloglucan-
CC specific endo-b-1,4-glucanase) enzyme derived from the Geotrichum genus
CC and having the ability to hydrolyse the beta-1,4-glycosidic linkage of a
CC glucose residue, but not that of a xylose residue contained in a
CC xyloglucan. The polypeptide of the invention may be useful for
CC elucidating the structure or function of a xyloglucan having an important
CC role in the differentiation of a plant cell. The current sequence is that
CC of the Geotrichum sp. M128 xyloglucanendohydrolase precursor DNA of the
CC invention.
XX
XX SQ Sequence 2755 BP; 555 A; 920 C; 836 G; 444 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Pred. No.: 6.55e-54 Length: 2755
XX Score: 979.00 Matches: 264
XX Percent Similarity: 48.43% Conservatve: 122
XX Best Local Similarity: 33.12% Mismatches: 279
XX Query Match: 24.26% Indels: 132
XX DB: 13 Gaps: 33
XX
XX US-09-917-376-3 (1-740) x ADR90295 (1-2755)
XX
XX 11 ValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGly 30
XX |||||
XX 283 GTGACGATCTCGGTGGCGGCTTATCTCGGCGCTCGTGGCACACCCGACGAGAAGGAC 342
XX |||||
XX 31 IleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTTPAspAlaAlaAsnGlyArg 50
XX |||||
XX

```

```

Db 343 CTGATCTACGCACGACAGACATTGGCGGCACGTACAGGTGGAACCGCGCCACAGTGGGAG 402
Qy 51 TrpIleProLeuLeuAspTrpVal-----GlyTrpAsnAsnTrpGlyTyr 65
Db 403 TGGAGAGCCATCAGGACTTTATCATCAACAGCGCTCGCGGCAACGGCGCAACCTG 462
Qy 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85
Db 463 CTCGCGCCGAGAGCATTGGCTCGACCCCAACCCAGACCGCTGTACCTCGCGCAA 522
Qy 86 GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGln 105
Db 523 GCGCACTATGTG---CAGTGGGACCG---TGGGCGGCTTCTCTCGTCTCGGACGACCG 576
Qy 106 GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArg 125
Db 577 GCGAAGAGCTTCAGCAGCAGTACCGCTCGGCCCTGTCCTCCATCGGCGCCCAACACATCGGGCGC 636
Qy 126 GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 145
Db 637 AATGGCGGAGCGCTCGCGGTCAACCGCACTGGACCGAGAGCTGTGTTTGGCTCG 696
Qy 146 ProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 165
Db 697 CGCAGC---CAGGGCTGTGGCGCTCCACTGACCGCGCTCAGACGCTGCTCGGCATGAAC 753
Qy 166 AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSer 185
Db 754 CAGTCTCCCGACAGCTCGACGTAC-----GGAATC----- 783
Qy 186 AspileGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGlnAla 205
Db 784 -----GGCATCATCTCGGTCACTTCGACCCCAAGAACGTCGCGCAGCATACGTC 834
Qy 206 SerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArgAsp 225
Db 835 GCTTCGCACGCGCTCGCGGGCTGTGG-----GTACAGTGG-----GAC 873
Qy 226 GlyGlyAlaThrTrpGlnAlaValProGlyAlaProThr----- 238
Db 874 GCGCGCGCAACTGGTCCAGGTGGCGCGCCAGCTACGAGTGTGTCAGCTGACAGCAAAAG 933
Qy 239 -----GlyPheIleProHisLysGly 245
Db 934 AGCATTGTGCTGCTCGCGCACCGCATCCAGAGCTCGGGGCCCTCCCGATCAAGATC 993
Qy 246 ValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyr 265
Db 994 GCACCTGGGCAAG---AACGGCAGGCTGTATACCTACTCGGACGACCGCGGCCCTGG 1050
Qy 266 AspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArgIle 285
Db 1051 GCGCTCTCTACGGCGAGGTCTGTGTCCTACGATCCCAACGCAACGCAACGCAACATC 1110
Qy 286 SerProValProSerThrAspThrAlaAsnAspTyr----- 297
Db 1111 ACC-----CCCTCGCGGCGGCGCCCAACAGTACCTCTGCCCGCCACTGGCAACAAAAG 1164
Qy 298 -----PheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMet 314
Db 1165 GTCGTTCCCGGAGGCTGGAAAGGTATCTCGGTCGGT-----AACGGCAGACCGTCGTC 1218
Qy 315 ValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGly 334
Db 1219 GTGTCGAGCTCGACGCAAAACGCGCAAGACTCA---GTGTACTCTCTCGCGCAGCGCGC 1275
Qy 335 AlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrVal 354
Db 1276 AACTCTGGAAAGGACTCGGCAAGCTGACACACGCGCGCGCGCGCGCGGCGCAACTCGCAG 1335
Qy 355 LeuAspileSerAlaGlu-----ProTrpLeuThrPheGlyVal 367
Db 1336 AAGGAGTCGAGCCCAAGCTGGCAACGCGCACTCCGCTCCCTTGGTCTCTGTTCCAGAAC 1395

```







```
DB: 10 Gaps: 52
US-09-917-376-3 (1-740) x ADD42054 (1-5698)
QY 7 ThrTrpSerAsnValAlaIle---GlyGlyGlyPheValAspGlyIleValPheAsn 25
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 242 TCATGGAAGAACGTCAAGCTCGGCGCGCGCGCTTCGTCCCGGCATCATCTTCCAA 301
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 26 -----GluGlyAlaPro----- 29
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 302 AGTACCTTTCTGCACTTCGAGCCGCCGCCGCCGCGCGAGACGAGCGCGGTAGTAGAGTT 361
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 30 -----GlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrp--- 44
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 362 CCCAAGACAAGGGGTAGCATATGCCAGACAGATATTTGGCGGCTGTACCGCTCAAA 421
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 45 -----AspAla 46
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 422 GGGTCTCTGTTTCGCATCGTATACGTCTGTCTATAACCCCGCAGCATGGCGAGTTC 481
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 47 AlaAsnGlyArgTrpIleProLeuLeuAspTrpVal-----GlyTrp----- 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 482 CCGCAGCATCATGAGCCCGCTACGATGGGATTGCTGATAATGCGCGGTGGCACAAG 541
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 -----AsnAsnTrpGly-----Tyr 65
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 542 CGGCTGCTGAGTACCTGGCGGCGAGTGCCTACCTAACGACTATTACGCCGCGCTGTT 601
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnIleValTrpAlaVal 85
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 602 TGGGCGCATCGAGCTGTTGGCTTCATCCGCGAGGACGATCAAAAGGTGTATGCCGCA 661
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 85 ----- 85
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 662 ACCCGTAGTTCGCAACACGCGAAGTACGAGTCTGCTAGTCTTCCACATACGCGCTCAC 721
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 86 GlyMetTyrThrAsnSerTrpAsp----- 93
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 722 GGCATGTATACAGACAGCTGTGTC-TGTGATGTCTCTCAGATCTAGACCTATGATTTGGAG 780
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 93 ----- 93
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 781 GCGGTACATATGCTTGTGCGCAGACACTACAGAGAGTCTAGATCTGGATACTAACCTGC 840
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 94 -----ProAsnAspGlyAlaIleLeuArgSerSerAspGln 105
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 841 CTGACATTGGCCATATATAGGATCCGAGTAAATGGAGCCATCATTTCTGCTCGTCAGAC 900
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 105 ----- 105
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 901 GACTGTAAACCGGTATATATCCCTAGGCTCATTAACCTCGGTAGTAAGCGAGCAGTCTGG 960
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 106 GlyAlaThrTrpGlnIleThrProLeuProPheLeuLysLeuGlyGlyAsnMetProGlyArg 125
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 961 GCGCAACGTGTCTCTCACCAACTTGCCTTCAAGTCGGGGGTAAACATGTCAGGACGC 1020
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 125 ----- 125
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1021 CCGCGTTCACCGAGAGTGGTTGAACGGGAAGTTTCAGCCGCCCATTTGTACGGTCTCG 1080
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 126 GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 145
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1081 GGAGCGGAGAGCGTGTGCTGTCTGATCCGGCCAACTCCAGCACTCATCTACTTGTGTCT 1140
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 146 Pro----- 146
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1141 CTTCCGCTCTCGCAGCCGACAGCTAGCCCGGTTGAGTTGTAGTAGATGAACCAACA 1200
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 147 ---SerGlyLeuGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMet--- 164
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1201 CGCTCAGGAACCGCTCTGGAGTCTACGGACGGCGGTGACCTTTTCCAAAGTCTCG 1260
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 165 -----Thr 165
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```



Db 2341 AAAGCAGCTGGCTGAGACC--CTGTTGTACCTCGGGCTAGACCCGACCCGCTCGATAGG 2399  
QY 362 -----Trrp 362  
Db 2400 CACTGAGACCTATTACTACAGCATCTCAGTGAGTCACCTTTAAACGATCCGATGCGAATGG 2459  
QY 363 LeuThrPhe-----GlyValGlnProAsn----- 370  
Db 2460 CTGACTCTGGGATAATGATGTGTGATGATCCTCAGTGAGAAATGTGATGGCTACGCTTACC 2519  
QY 371 -----ProProValPro----- 374  
Db 2520 GACTGACTCTGCTTCAAGACTCCCAAGACCGTGGATCAAGAACAACTTTATCGATGTG 2579  
QY 374 ----- 374  
Db 2580 ATGACTGAGACGAAGTTCTGAGGGTTTCTGTGGCACCTAGTTCTTGTGAAATAGCTACAC 2639  
QY 375 -----SerPro-----LysLeuGlyTrrpMetAspGluAlaMet 385  
Db 2640 TCGACGAGTCAACCGTCCGATCCGATGCTCATCAAGCCGCTCGGCTGGATGATGATGCTCTC 2699  
QY 386 Ala----- 386  
Db 2700 GGCTCGCTCAGTGGCAGGCTACCAAGTAGTTTCGGGAGCCGACCTACTAATCAGAGAG 2759  
QY 387 ---IleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAla 405  
Db 2760 CAGATTGACCAACAGCAGCAACCACTGGCTCTACGGCACCGAATGACAACTCTTTGGC 2819  
QY 406 Thr----- 406  
Db 2820 GTCTAACTGGTGTGCTGCTGTGGTGACCGAGATGCGCGTGCCTTACTGTTAGAAACCG 2879  
QY 407 ---AsnAspLeuThrLysTrrpAspSerGlyGlyGlnIleHisIleAlaProMetValLys 425  
Db 2880 CCGCAGCATCTCACCACATGGGACACGGCCACATGTGTCAATCCATCACTGSCAGAC 2939  
QY 426 Gly----- 426  
Db 2940 GCGTGTAGAGTGTGTGACCTGTGCGGGTGTACACAGTTAGTTAGTACCGCTCTG 2999  
QY 427 ---LeuGluGluThrAlaValAsnAspLeuIleSerProProSerGlyAlaProLeuIle 445  
Db 3000 CGCATCGAGGAATCTCTCGTCCAGGACCTGGCTCTGCACCCGCGGAGCGAGTATTG 3059  
QY 446 Ser----- 446  
Db 3060 GCGTAGCTCCTTAAGAGGAGGTCTCTGGACCGGAGACGTGGGCGCGCTTCGCTCGATAC 3119  
QY 447 ---AlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaVal----- 462  
Db 3120 CCGGAGTCCGAGACGACAAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGACATCG 3179  
QY 462 ----- 462  
Db 3180 CCGGCTCAGCCTCTGCTGTTGCCGAAGTGGAAACGGTCTGCTTTGTGGAGCCCTGTAGC 3239  
QY 463 ProSerThrIlePheThrSerProValPheThrGlyThrSerValAspTyrAla--- 481  
Db 3240 GCGCAGCGGTCTGGGCAACGCCCATATGGGCCACCTCGACGAGCGTCACTACGCCGG 3299  
QY 482 -----Glu 482  
Db 3300 AGGTCTGCCACACCGTTGGGTGTATCCCGGTGGAGTGTCTCGCAGTGTATGTCGGCCC 3359  
QY 483 LeuAsnProSerIleIleValArgAlaGlySer----- 493  
Db 3360 TACTCGGTCAAGAGCGTCTCGCGTCGCGCAACACCGCGGACGCAACAGAGTGGCCATC 3419  
QY 494 -----PheAspProSerGlnProAsnAspArgHisValAlaPhe----- 507  
Db 3420 TTGAGCCAGTTCTCGCAGCAGGCGCAGCGGTGTGTGGCGCGGTGTGTTCACCCGGTAG 3479

QY 508 SerThrAspGlyGlyLysAsnTrrpPheGlnGlySerGluProGlyGlyValThrThrGly 527  
Db 3480 AGTCCGACGGGGCGGCGAGCATCGACTACGCGGCCGACACGTCATGAACGGC 3539  
QY 528 Gly----- 528  
Db 3540 GCGAGCTGCCCGCGCTGTCACCTCGTAGCTGATGCGCGGCTGTGCAGGTACTTTCGCG 3599  
QY 529 ---ThrValAlaAlaSerAlaAspGlySerArgPheValTrrpAlaProGlyAspProGly 547  
Db 3600 GCGACGGTGGCTATTTCGGCGGACGCGGACACGATCTCTGGTCGACCGCTCGTCCGCG 3659  
QY 548 -----GlnPro 549  
Db 3660 GGTGCCACCGATAGCCGGCTGCCCTGTGTAGGAGACAGCTGGCGGAGCAGCGCG 3719  
QY 550 ValValTyrAlaValGlyPheGlyAsnSerTrrpAlaAlaSerGlnGlyValProAlaAsn 569  
Db 3720 CTCGACGCTCG--CAGTTCCAGGGCAGCTTTGCCCTCCGCTCTCGAGCTCGCGCGCGC 3776  
QY 570 AlaGln----- 571  
Db 3777 GCGCAGCTCGCAGCGCTCAAGTCCGTCGAAACGGAGGACAGCTCGGACGGCGCGCG 3836  
QY 572 -----IleArgSerAspArgValAsnProLysThrPheTyrAlaAlaLeuSerAsnGlyThr 589  
Db 3837 CGGTCTATCGCTCGGACGAGCAAGACCAACAGCGTCTTCTACGCGGCTCGGATCGACC 3896  
QY 590 Phe----- 590  
Db 3897 TTTTAGTAGCGGAGCGCTGTTCTTCTGTTGTCGAGAAGATGCGGCGGAGCGCTAGCTGG 3956  
QY 591 ---TyrArgSerThrAspGlyValThrPheGlnProValAlaAlaGlyLeuProSer 609  
Db 3957 AAAAACGTCAGCAAGACACCGCGCAGCAGCTTC-----ACGCGCGGG---CCCAAG 4004  
QY 610 SerGlyAlaValGlyValMet----- 616  
Db 4005 GTGGGAGCGCAGGATGAGTGTCTGTCGCGCTGTCGAAAGTGGCGCGCGCGGTTTC 4064  
QY 617 -----PheHisAlaValProGlyLysGluGlyAspLeu 627  
Db 4065 GACCGTCCGCTCCCTCGATCCGGATATCGCTGCTCACCGACACCGCGGCGCAGCTTG 4124  
QY 628 TrpLeuAlaAlaSerSer----- 633  
Db 4125 TATGCTCGACCGCAGCGCTAGGCCCTATAGCAGAGTGGGCTGTGGCGCGCTGCAAC 4184  
QY 634 -----GlyLeuTyrHisSerThrAsnGlyGlySerSer----- 644  
Db 4185 ATACAGAGTGGCTGTCTCGGCATATTCCGCTCCACAGACTCGGGCAGCAGCTTTGGCCAA 4244  
QY 644 ----- 644  
Db 4245 GTCTCCACCGCTGAAGCCGTATAGCGGAGGTGTCTGAGCCGTGTGGAACCGGTT 4304  
QY 645 ---TrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSer 662  
Db 4305 CAGAGTGGCGGAGCTCCAAACACTACAGATCGCCCTGGGTGGGCTCAGGCTCGAAC 4364  
QY 663 -----AlaProGlySerSerTyr 668  
Db 4365 TGGAACTGTATGCTGTTGTGATGTTCTAGCGGAGCCACACCGAGTCCGAGCTTG 4424  
QY 669 ProAlaValPheValValGlyThrIleGlyGlyValThrGlyAla-----TyrArgSer 686  
Db 4425 ACCTTGGACATACGATCCGACCC-----GGCCGCTCAGGGCTCGCCTCTACGCCAGT 4478  
QY 687 AspAspCysGlyThrThrTrrpVal----- 694  
Db 4479 GGAGACAGCGCGCCTCTCTGGAAGCCGTGGCGGCGAGTCCCGAGCGGAGATGCGGTCA 4538



```

Qy 695 -----LeulleAsnAspAspGlnHisGlnTyrGly----- 704
Db 4539 CCTCTGTCGCGCGGAGGACCTCGACATCCAGGGCTCCAGGGCTTCGGGCTCCATCGAC 4598
Qy 704 ----- 704
Db 4599 AGCACCAAGGTGCGCGGCAGCGGCTGTAGGTCCCGAGGGTCCCGAAGCCGAGGTAGCTG 4658
Qy 705 AsnTyrGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThr 724
Db 4659 TCGTGTGTTCCAGCGCGCGTCCGCGCAGCACCAGCGCGG-----CAAGTCTACGTGGGCACC 4712
Qy 725 AsnGlyArgGlyIleValTyrGlyAspIleGlyGlyAlaPro 738
Db 4713 AACGGCCGGGGCGTCTTTTACGCTCAGGCGTCTGTGGCGGCC 4754

```

Search completed: October 5, 2005, 09:20:44  
Job time : 1225.43 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 5, 2005, 07:43:21 ; Search time 7115.69 Seconds  
(without alignments)  
3958.515 Million cell updates/sec

Title: US-09-917-376-3  
Perfect score: 4036  
Sequence: 1 ATTQPTWSNVAIGGGFVD.....YGTNGRGIVYDGGAPSG 740

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0  
Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-O=/cpn2.1/USPTO.spool/US09917376/runat\_04102005\_164328\_26911/app\_query.fasta\_1.2524  
-DB=EST\_QPWT-fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPECL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09917376 @CN 1 1 12009 @runat\_04102005\_164328\_26911 -NCPU=6 -ICPU=3  
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:\*

1: gb\_est1: \*  
2: gb\_est2: \*  
3: gb\_hcc: \*  
4: gb\_est3: \*  
5: gb\_est4: \*  
6: gb\_est5: \*  
7: gb\_est6: \*  
8: gb\_gss1: \*  
9: gb\_gss2: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	615.5	15.3	704	7	CF872104
2	615.5	15.3	761	6	CB902332
3	587.5	14.6	747	7	CF876916
4	587.5	14.6	814	6	CB905388
5	415.5	10.3	707	7	CF880713
6	415.5	10.3	782	6	CB907625
7	385.5	9.6	693	7	CF882065
8	361.5	9.0	751	7	CF868882
9	361.5	9.0	803	6	CB898982

10	338	8.4	738	7	CF866388
11	338	8.4	794	6	CB896354
12	314.5	7.8	546	2	BF072664
13	306.5	7.6	389	5	BU639045
14	286.5	7.1	713	7	CF875815
15	286.5	7.1	782	6	CB904767
16	282.5	7.0	719	7	CF881775
17	282.5	7.0	775	6	CB908435
18	278.5	6.9	929	7	CF885920
19	222	5.5	796	7	CF391910
20	183	4.5	3057	3	AK044947
21	180	4.5	4355	5	AY416870
22	171.5	4.2	296	5	BU641770
23	170	4.2	4631	3	AK082944
24	166.5	4.1	338	2	AW064456
25	153.5	3.8	3623	3	AK083848
26	153.5	3.8	3751	3	AK080277
27	153.5	3.8	4101	9	AY402312
28	151	3.7	4654	3	AK086725
29	150.5	3.7	5041	3	HSM806119
30	148.5	3.7	1178	8	CC203127
31	140	3.5	3782	3	AK041860
32	139.5	3.5	5367	9	AK084803
33	138.5	3.4	5370	9	CL976670
34	138	3.4	2847	9	CL982215
35	138	3.4	5073	9	AY407742
36	137.5	3.4	1886	3	AK009469
37	137	3.4	1581	3	BC029959
38	136.5	3.4	1569	9	AY407967
39	136.5	3.4	3200	3	HSM802592
40	136	3.4	1674	9	AY412937
41	134	3.3	426	5	BO751758
42	133.5	3.3	995	6	CD780482
43	133.5	3.3	2545	6	CD014011
44	132.5	3.3	769	6	CD788674
45	132.5	3.3	3006	9	CL982143

#### ALIGNMENTS

CF872104 704 bp mRNA linear EST 31-OCT-2003  
trico29xol3.b1 T.reesei mycelial culture, Version 6 October 2003  
Hypocrea jecorina cDNA clone trico29xol3, mRNA sequence.

CF872104 GI:38126786  
EST.  
Hypocrea jecorina (anamorph: Trichoderma reesei)  
Hypocrea jecorina  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.  
1 (bases 1 to 704)  
Diener,S.E., Dunn-Coleman,N., Foreman,P., Houfek,T.D.,  
Teunissen,P.J.M., van Solingen,P., Dankmeyer,L., Mitchell,T.K.,  
Ward,M. and Dean,R.A.  
Characterization of the protein processing and secretion pathways  
in a comprehensive set of expressed sequence tags from Trichoderma  
reesei

FEMS Microbiol. Lett. 230 (2), 275-282 (2004)

#### JOURNAL

#### COMMENT

Contact: Ralph A. Dean  
Fungal Genomics Laboratory  
North Carolina State University  
Campus Box 7251, Raleigh, NC 27695, USA  
Tel: 919-513-0020  
Fax: 919-513-0024  
Email: ralph\_dean@ncsu.edu

Seq primer: Lt-F1 primer.  
Location/Qualifiers  
1. .704

#### FEATURES

source  
/organism="Hypocrea jecorina"  
/mol\_type="mRNA"  
/strain="QM6a"



```
/db xref="taxon:51453"
/clone="trico29xol3"
/dev_stage="mycelia"
/clone_lib="T.reesei mycelial culture, Version 6 October
2003"
/notes="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial
culture grown from 24 hrs to 6 days with varying Carbon
and Nitrogen sources and concentrations."

ORIGIN

Alignment Scores: 6.93e-46 Length: 704
Pred. No.: 615.50 Matches: 118
Score: 615.50 Matches: 118
Percent Similarity: 75.61% Conservative: 37
Best Local Similarity: 57.56% Mismatches: 43
Query Match: 15.25% Indels: 9
DB: 7 Gaps: 4

US-09-917-376-3 (1-740) x CP872104 (1-704)

Qy 6 TyrThrTrpSerAsnValAlaIle---GlyGlyGlyGlyPheValAspGlyIleValPhe 24
Db 98 TTTTCATGGAAGAACGCTCAAGCTCGCGCGCGCGCGCTTCGTCGCCGCGCATCATCTTC 157
Qy 25 AsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArg--- 43
Db 158 CATCCCAAGCAAAAGCGGTAGCATATGCACGACAGATATTTGGCGGCTGTACCGCCTC 217
Qy 44 TrpAspAlaAlaAsnGlyArgTyrIlePro-----LeuLeuAspTyrValGlyTyrAsn 61
Db 218 TTCTCCGCGACTCATGACCGCC-GTCACGGATGGATGTCTGATATATGCGCGGTGGCAC 276
Qy 62 AsnTrpGlyTyrAsnGlyValValSerIleAlaAspProIleAsnThrAsnIleVal 81
Db 277 AACTGG-----GGCATCGACGCTGTTCGCTTGATCCGACGACGATCAAAAGGTG 327
Qy 82 TrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyValAlaIleLeuArg 101
Db 328 TATCCCGAGTCGCGATGTATACGACAGCTGGGATCCGATATGAGTATGAGCCATCATTCGC 387
Qy 102 SerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsn 121
Db 388 TCGTCAGACCGCGCGCAACGCTGCTCTCCACCACTTGCCCTTCAAAAGTGGGGGTAAAC 447
Qy 122 MetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeu 141
Db 448 ATGCCAGACCGGAGCGGAGAGCGTCTGGCTGTGCGATCCGCGGCAACTCCCAACATCATC 507
Qy 142 TyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyValAlaThrTrp 161
Db 508 TACTTTGGTCTCGCTCAGGAACAGCGCTCTGGAAGTCTACGACGCGCGGTGACCTTT 567
Qy 162 SerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThr 181
Db 568 TCCAAGTCTCGTTCACGCAACTGGGAGTACATCCCGACCCCGAGTATTCACAC 627
Qy 182 GlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSer 201
Db 628 GGCTACCACACGATCAGACGAGTCTATGGGTACGTTTACGATC-ACCAGCAGCAGC 686
Qy 202 LeuGlyGlnAlaSer 206
Db 687 ACCGGNGAGCCACG 701

RESULT 2
CB902332
LOCUS CB902332
DEFINITION trico29xol3 T.reesei mycelial culture, Version 3 april Hypocrea
jecorina cDNA clone trico29xol3, mRNA sequence.
ACCESSION CB902332
VERSION CB902332.1 GI:30116990
KEYWORDS EST.
SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)
```

ORGANISM Hypocrea jecorina  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.

REFERENCE 1 (bases 1 to 761)  
Foreman, P.K., Brown, D.E., Dankmeyer, L., Dean, R., Diener, S.,  
Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., England, G.J.,  
Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchinson, C.,  
Olivares, H.A., Teunissen, P.J., Yao, J. and Ward, M.  
Transcriptional regulation of biomass-degrading enzymes in the  
filamentous fungus *Trichoderma reesei*  
J. Biol. Chem. 278 (34), 31988-31997 (2003)

JOURNAL MEDLINE  
22803314  
PUBMED 12788920  
COMMENT Contact: Pamela K. Foreman  
Genencor Intl.  
925 Page Mill Road, Palo Alto, CA 94304, USA  
Tel: (650) 846-7635  
Fax: (650) 621-7817  
Email: Pforeman@genencor.com  
Seq primer: LT-F1 primer.

FEATURES  
source  
1..761  
Location/Qualifiers  
/organism="Hypocrea jecorina"  
/mol\_type="mRNA"  
/strain="QM6a"  
/db\_xref="taxon:51453"  
/clone="trico29xol3"  
/dev\_stage="mycelia"  
/clone\_lib="T.reesei mycelial culture, Version 3 april"  
/note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial  
culture grown from 24 hrs to 6 days with varying Carbon  
and Nitrogen sources and concentrations."

ORIGIN

Alignment Scores: 7.81e-46 Length: 761  
Pred. No.: 615.50 Matches: 118  
Score: 615.50 Matches: 118  
Percent Similarity: 75.61% Conservative: 37  
Best Local Similarity: 57.56% Mismatches: 43  
Query Match: 15.25% Indels: 9  
DB: 6 Gaps: 4

US-09-917-376-3 (1-740) x CB902332 (1-761)

Qy 6 TyrThrTrpSerAsnValAlaIle---GlyGlyGlyGlyPheValAspGlyIleValPhe 24  
Db 155 TTTTCATGGAAGAACGCTCAAGCTCGCGCGCGCGCTTCGTCGCCGCGCATCATCTTC 214  
Qy 25 AsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArg--- 43  
Db 215 CATCCCAAGCAAAAGCGGTAGCATATGCACGACAGATATTTGGCGGCTGTACCGCCTC 274  
Qy 44 TrpAspAlaAlaAsnGlyArgTyrIlePro-----LeuLeuAspTyrValGlyTyrAsn 61  
Db 275 TTCTCCGCGACTCATGACCGCC-GTCACGGATGGATGTCTGATATATGCGCGCTGTACCGCCTC 333  
Qy 62 AsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnIleVal 81  
Db 334 AACTGG-----GGCATCGACGCTGTTCGCTTGATCCGACGACGATCAAAAGGTG 384  
Qy 82 TrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArg 101  
Db 385 TATGCCGCGATCGGCATGTATACGACAGCTGGATCCGATGAGTATGAGCCATCATTCGC 444  
Qy 102 SerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsn 121  
Db 445 TCGTCAGACCGCGCGCAACGCTGGTCTTCACCAACTTGCCCTTCAAGTGGGGGTAAAC 504  
Qy 122 MetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeu 141  
Db 505 ATGCCAGACCGGAGCGGAGAGCGTCTGGCTGTCTGATCCGCGCAACTCCCAACATCATC 564  
Qy 142 TyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThrTrp 161











Db 438 ACGTGGAGCATGCTAGCGCGCCGACACACGTCCTCCATGAACGGCGGACACGCTGCGCTATTTCG 497  
 Qy 534 AlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTrpAla 553  
 Db 498 GCCGACGGCGACACGATCCTCTGTCGACCGGCTCGTCCGGC-----GTGACGGCG 548  
 Qy 554 ValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArg 573  
 Db 549 TCGAGTTCCAGGCGAGCTTTGCTCCGCTCGAGCTCGCCGCGCGGCGCGCTCATCGCC 608  
 Qy 574 SerAspArgValAsnProLysThrPheTyAlaLeuSerAsnGlyThrPheTyArgSer 593  
 Db 609 TCGGACAAAGAGACCAACAGCGCTCTTACGCGCGCTCGGATCGACCTTTTACGTCAGC 668  
 Qy 594 ThrAspGlyGlyValThrPhe 600  
 Db 669 AAGGACACCGCGCAGCAGCTTC 689

RESULT 6  
 CB907625 782 bp mRNA linear EST 02-JUL-2003  
 LOCUS tric082xn19 T.reesei mycelial culture, Version 3 april Hypocrea  
 DEFINITION jecorina cDNA clone tric082xn19, mRNA sequence.

ACCESSION CB907625  
 VERSION CB907625.1 GI:30122283  
 KEYWORDS EST.  
 SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)  
 ORGANISM Hypocrea jecorina  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 Hypocremycetidae; Hypocreales; Hypocreaceae; Hypocrea.

REFERENCE 1 (bases 1 to 782)  
 Foreman,P.K., Brown,D.E., Dankmeyer,L., Dean,R., Diener,S.,  
 Dunn-Coleman,N.S., Goedegebuur,F., Houfek,T.D., England,G.J.,  
 Kelley,A.S., Meerman,H.J., Mitchell,T., Mitchinson,C.,  
 Olivares,H.A., Teunissen,P.J., Yao,J. and Ward,M.  
 Transcriptional regulation of biomass-degrading enzymes in the  
 filamentous fungus Trichoderma reesei

JOURNAL J. Biol. Chem. 278 (34), 31988-31997 (2003)  
 MEDLINE 22803314  
 PUBMED 12788920  
 COMMENT Contact: Pamela K. Foreman  
 Genencor Intl.  
 925 Page Mill Road, Palo Alto, CA 94304, USA  
 Tel: (650) 846-7635  
 Fax: (650) 621-7817  
 Email: pforeman@genencor.com  
 Seq primer: LN-F1 primer.

FEATURES  
 source  
 1..782  
 /organism="Hypocrea jecorina"  
 /mol\_type="mRNA"  
 /strain="QM6a"  
 /db\_xref="taxon:51453"  
 /clone="tric082xn19"  
 /dev\_stage="mycelia"  
 /clone\_lib="T.reesei mycelial culture, Version 3 april"  
 /note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial  
 culture grown from 24 hrs to 6 days with varying Carbon  
 and Nitrogen sources and concentrations."

ORIGIN  
 Alignment Scores:  
 Pred. No.: 1,85E-27 Length: 782  
 Score: 415.50 Matches: 92  
 Percent Similarity: 55.51% Conservative: 34  
 Best Local Similarity: 40.53% Indels: 90  
 Query Match: 10.29% Gaps: 3  
 DB: 6  
 US-09-917-376-3 (1-740) x CB907625 (1-782)

Qy 377 LysLeuGlyTrpMetAspGlnAlaMetAlaAlaAspProPheAsnSerAspArgMetLeu 396  
 ::

Db 108 CCGCTCGGCTGGATGATTGAGTCTCTCGAGATTGACCCACGACGACACCACTGGCTC 167  
 Qy 397 TyTGlyThrGlyValThrLeuTyAlaThrAsnAspLeuThrLysTrpAspSerGlyGly 416  
 Db 168 TAGCGACCGGATCAATCAATCTTTGGGGCCACGATCTCACCAACTGGGACACGGCCAC 227  
 Qy 417 GlnIleHisIleAlaProMetValLysGlyLeuGlnGluThrAlaValAsnAspLeuIle 436  
 Db 228 AATGTGTCAATCAATCACTGGCAGACGGCATCGAGGAATTCCTCCGTCCAGGACCTGGCC 287  
 Qy 437 SerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHis 456  
 Db 288 TCTGCACCCGGCGAAGCGAGCTATTGGCCGCGAGTCGAGACGACCAACGGCTTCACCTTT 347  
 Qy 457 Ala-----AspValThrAlaValProSerThrIlePheThrSerProValPheThr 473  
 Db 348 GCCAGCAGAAACGACCTCGGGACATCGCCGACACGGTCTGGCAGACGCCACCATGGGCC 407  
 Qy 474 ThrGlyThrSerValAspTyAlaGluLeuAsnProSerIleIleValArgAlaGlySer 493  
 Db 408 ACTCGACGAGCGTGCAGTACGCGGGAACCTCGTCAAGAGCGTCTGTCGCGTCGGCAAC 467  
 Qy 494 PheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLys 513  
 Db 468 ACCGCGCGCAGCAACAG-----GTGCGCATCTCTGTCGAGCGCGCGCGCGCG 512  
 Qy 514 AsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSer 533  
 Db 513 ACGTGGAGCATGCTAGCGCGCGGACAGCTCATGACCGCGGACGCGTGGCGCTATTCG 572  
 Qy 534 AlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTrpAla 553  
 Db 573 GCGCAGCGGCGACGATCTCTGTCGACCGCTCTGTCGCGC-----GTGACGGC 623  
 Qy 554 ValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArg 573  
 Db 624 TCGAGTTCCAGGCGCAGCTTTGCTCCTCGTCTCGAGCTGCGCGCGCGCGCTCATGCC 683  
 Qy 574 SerAspArgValAsnProLysThrPheTyAlaLeuSerAsnGlyThrPheTyArgSer 593  
 Db 684 TCGGACAAAGAGACCAACAGCGCTCTTACGCGCGCTCGGATCGACCTTTTTCGTCAGC 743  
 Qy 594 ThrAspGlyGlyValThrPhe 600  
 Db 744 AAGGACACCGCGCAGCAGCTTC 764

RESULT 7  
 CF882065 693 bp mRNA linear EST 31-OCT-2003  
 LOCUS tric029xol3.b11 T.reesei mycelial culture, Version 6 October 2003  
 DEFINITION Hypocrea jecorina cDNA clone tric029xol3, mRNA sequence.

ACCESSION CF882065  
 VERSION CF882065.1 GI:38136747  
 KEYWORDS EST.  
 SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)  
 ORGANISM Hypocrea jecorina  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.  
 1 (bases 1 to 693)

REFERENCE 1  
 Diener,S.E., Dunn-Coleman,N., Foreman,P., Houfek,T.D.,  
 Teunissen,P.J.M., van Solingen,P., Dankmeyer,L., Mitchell,T.K.,  
 Ward,M. and Dean,R.A.  
 Characterization of the protein processing and secretion pathways  
 in a comprehensive set of expressed sequence tags from Trichoderma  
 reesei

JOURNAL FEBS Microbiol. Lett. 230 (2), 275-282 (2004)  
 COMMENT Contact: Ralph A. Dean  
 Fungal Genomics Laboratory  
 North Carolina State University  
 Campus Box 7251, Raleigh, NC 27695, USA  
 Tel: 919-513-0020  
 Fax: 919-513-0024  
 Email: ralph\_dean@ncsu.edu







QY 584 AlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProVal 603  
 Db 531 GCCGCTCCGATCGACCTTTTACGTACAGGACACCGGACAGCTTC-----581  
 QY 604 AlaAlaGlyLeuProSerSerGlyAlaValGlyValMet-----PheHisAlaValPro 621  
 Db 582 ACACCGCGGG---CCCAAGCTGGCAGCGCAGGAGATCCGGATATCGTCTCACCCG 638  
 QY 622 GlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGly 641  
 Db 639 ACCACCGCGGACGCTTGATGTCGACCGACGTCGATATCGCTCCACAGACTCG 698  
 QY 642 GlySerSerTrpSerAlaIleThr 649  
 Db 699 GGCACGACCTTTGGCCAAAGTCTCC 722

RESULT 9  
 CB898982  
 LOCUS  
 DEFINITION  
 trico16xm14 T. reesei mycelial culture, Version 3 april Hypocrea  
 jecorina cDNA clone trico16xm14, mRNA sequence.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE

ORGANISM  
 Hypocrea jecorina (anamorph: Trichoderma reesei)

REFERENCE  
 AUTHORS  
 Foreman, P.K., Brown, D.E., Dankmeyer, L., Dean, R., Diener, S.,  
 Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., England, G.J.,  
 Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchinson, C.,  
 Olivares, H.A., Teunissen, P.J., Yao, J. and Ward, M.

TITLE  
 Transcriptional regulation of biomass-degrading enzymes in the  
 filamentous fungus Trichoderma reesei  
 J. Biol. Chem. 278 (34), 31988-31997 (2003)

JOURNAL  
 MEDLINE  
 PUBMED  
 COMMENT  
 Contact: Pamela K. Foreman  
 Genencor Intl.  
 925 Page Mill Road, Palo Alto, CA 94304, USA  
 Tel: (650) 846-7635  
 Fax: (650) 621-7817  
 Email: pforeman@genencor.com  
 Seq primer: Lr-F1 primer.

## FEATURES

source  
 1..803  
 /organism="Hypocrea jecorina"  
 /mol\_type="mRNA"  
 /strain="QM6a"  
 /db\_xref="taxon:51453"  
 /clone="trico16xm14"  
 /dev\_stage="mycelia"  
 /clone\_lib="T. reesei mycelial culture, Version 3 april"  
 /note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial  
 culture grown from 24 hrs to 6 days with varying Carbon  
 and Nitrogen sources and concentrations."

## ORIGIN

Alignment Scores:  
 Pred. No.: 1,74e-22 Length: 803  
 Score: 361.50 Matches: 88  
 Percent Similarity: 51.61% Conservative: 40  
 Best Local Similarity: 35.48% Mismatches: 103  
 Query Match: 8.96% Indels: 17  
 DB: 6 Gaps: 6

US-09-917-376-3 (1-740) x CB898982 (1-803)

QY 407 AsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGly 426  
 Db 67 CAGGATTTCCACCACTGGGACACCGCCCAATGTGTCAATCCATCATCTGGCAGCGGC 126

QY 427 LeuGluGluThrAlaValAsnAspLeuIleSerProProSerGlyAlaProLeuIleSer 446  
 Db 127 ATCCAGGAATTCTCCGTCCAGGACCTTGCGCTCTGCACCCGCGGAAGAGCTATTGGCC 186  
 QY 447 AlaLeuGlyAspLeuGlyGlyPheThrHisAla-----AspValThrAlaValPro 463  
 Db 187 GCAGTCGGAGACACAAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGGACATCGCCG 246  
 QY 464 SerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeu 483  
 Db 247 CAGACGGTCTGGCAACGCCACATGGCCACCTCGACGAGCGTGCACCTACGCGGGGAAC 306  
 QY 484 AsnProSerIleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArg 503  
 Db 307 TCGGTCAAGAGCGTCTCGCTCGCAACACCGCGCGCACCAACAG-----354  
 QY 504 HisValAlaPheSerThrAspGlyLysAsnTrpPheGlnGlySerGluProGlyGly 523  
 Db 355 ---GTGGCCATCTCGTCCGACGCGCGCGAGCTGGAGCATCGACTACGCGCGGACACG 411  
 QY 524 ValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTtpAlaPro 543  
 Db 412 TCCATGACCGCGGCACGCTGTCCTATTTCGCGCGCGGCGACACGATCTCTGTGTCGACC 471  
 QY 544 GlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSer 563  
 Db 472 GCTCTGCTCCGCG-----GTGCAGCGCTCGCAGTTCAGGGCAGCTTGCCTCCGTC 522  
 QY 564 GlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyr 583  
 Db 523 TCGAGCTCGCCCGCGCGCGCGTCACTCGCTCGCAAGAGAACCAACAGCGCTTCTTAC 582  
 QY 584 AlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProVal 603  
 Db 583 GCGCGCTCGGATCGACCTTTTACGTGACGACGACGCGGCGACGCTTC-----633  
 QY 604 AlaAlaGlyLeuProSerSerGlyAlaValGlyValMet-----PheHisAlaValPro 621  
 Db 634 ACACCGCGGG---CCCAAGCTGGCAGCGCAGGAGCATCCGGATATCGCTGCACCCG 690  
 QY 622 GlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGly 641  
 Db 691 ACCACCGCGGACGCTTGATGTCGACCGACGTCGCGCATATTCGCTCCACAGACTCG 750  
 QY 642 GlySerSerTrpSerAlaIleThr 649  
 Db 751 GGCACGACCTTTGGCCAAAGTCTCC 774

## RESULT 10

CF866388  
 LOCUS  
 DEFINITION  
 trico06xe08.b1 T. reesei mycelial culture, Version 6 October 2003  
 Hypocrea jecorina cDNA clone trico06xe08, mRNA sequence.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## CF866388

## trico06xe08.b1

## T. reesei

## mycelial culture,

## Version 6 October 2003

## Hypocrea jecorina

## cDNA clone

## trico06xe08,

## mRNA sequence.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## CF866388

## trico06xe08.b1

## T. reesei

## mycelial culture,

## Version 6 October 2003

## Hypocrea jecorina

## cDNA clone

## trico06xe08,

## mRNA sequence.



Tel: 919-513-0020  
 Fax: 919-513-0024  
 Email: ralph.dean@ncsu.edu  
 Seq primer: LT-F1 primer

# FEATURES

source  
 1..738  
 /organism="Hypocrea jecorina"  
 /mol\_type="mRNA"  
 /strain="QM6a"  
 /db\_xref="taxon:51453"  
 /clone="tric006xe08"  
 /dev\_stage="mycelia"  
 /clone\_lib="T.reesei mycelial culture, Version 6 October 2003"  
 /note="Vector: PREP3Y; Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

## ORIGIN

Alignment Scores:  
 Pred. No.: 2,19e-20 Length: 738  
 Score: 338.00 Matches: 89  
 Percent Similarity: 51.97% Conservative: 43  
 Best Local Similarity: 35.04% Mismatches: 102  
 Query Match: 8.37% Indels: 20  
 DB: 7 Gaps: 9

US-09-917-376-3 (1-740) x CF866388 (1-738)

QY 422 ProMetValLys--GlyLeuGluThrAlaValAspLeuSerProProSer 440  
 Db 2 CCACGGTCCGAGCGCATCGAGAAATTCCTCGTCCAGGACCTGGCTCTGCACCGGC 61  
 QY 441 GlyAlaProLeuSerAlaLeuGlyAspLeuGlyGlyPheThrHisAla----- 457  
 Db 62 GGAAGCGAGCTATTGCGCGCAGTCGGAGACGACAAACGGCTTCCACCTTTGCCAGCAAAAC 121  
 QY 458 AspValThrAlaValProSerThrPheThrSerProValPheThrThrGlyThrSer 477  
 Db 122 GACCTCGGAGATCCCGCGAGACGGTCTGGGACGCGCCACATGGCCACCTCGACGAGC 181  
 QY 478 ValAspTyrAlaGluLeuAsnProSerIleValArgAlaGlySerPheAspProSer 497  
 Db 182 GTCGACTACCGCGGAATCTCGTCAAGCGCTCGTCCGGTCCGCAACACCGCCGCGCAG 241  
 QY 498 SerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGln 517  
 Db 242 CAACAG-----GTGGCCATCTCGTCGACGGCGCGCGCGACGTGGAGCATC 286  
 QY 518 GlySerGluProGlyGlyValThrGlyGlyThrValAlaAlaSerAlaAspGlySer 537  
 Db 287 GACTACGGCGCGCACACGTCCTATGAACCGCGCGACGGTGGCTATTCGGCCGACGCGCAG 346  
 QY 538 ArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGly 557  
 Db 347 AGCATCTCTGTCTGACCGCCCTCGTCCGCG-----GTGACGGCTCGCAGTTCAG 397  
 QY 558 AsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgVal 577  
 Db 398 GGCAGCTTGCTCTCGTCTCGAGCCTCGCGCGCGCGCGCGCTCATCGCTCGGACCAAGAAG 457  
 QY 578 AsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGly 597  
 Db 458 ACCAAGCGCTTCTTACGCGCGCTCCGGATCGACCTTTTACGTGACGAAGACACCGCGC 517  
 QY 598 ValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMet--- 616  
 Db 518 AGCAGCTTC-----ACGCGCGGG-----CCCAAGCTCGGCGCGCGCGAGCGATCCGG 565  
 QY 617 ---PheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyGly 635  
 Db 566 GATATCGCTGCTCACCGGACACCGCGGCGCGCGCTGTGTATGTCTCGACCGACGCTCGGCATA 625

QY 636 TyrHisSerThrAsnGlyGlySerSerTrpSerAlaAlaIle---ThrGlyValSerSerAla 654  
 Db 626 TTCCGCTCCACAGCTCGGCGACGACCTTTGGCCAAGTCTCCACCGCCCTGCACCAACACC 685  
 QY 655 ValAsnValGlyPheGlyLysSerAlaProGlyLysSerTyr 668  
 Db 686 TACCAGATCGCCCTCGGTGTGGGCTCA---GGCTCGAAGTGG 724

## RESULT 11

CB896354 794 bp mRNA linear EST 02-JUL-2003  
 LOCUS CB896354  
 DEFINITION jecorina cDNA clone tric006xe08, mRNA sequence.

ACCESSION CB896354

VERSION CB896354.1 GI:30111012

KEYWORDS EST.

SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)

ORGANISM Hypocrea jecorina

REFERENCE 1 (bases 1 to 794)  
 AUTHORS Foreman,P.K., Brown,D.E., Dankmeyer,L., Dean,R., Diener,S.,  
 Dunn-Coleman,N.S., Goedegebuur,F., Houfek,T.D., England,G.J.,  
 Kelley,A.S., Meerman,H.J., Mitchell,T., Mitchinson,C.,  
 Olivares,H.A., Teunissen,P.J., Yao,J. and Ward,M.

TITLE Transcriptional regulation of biomass-degrading enzymes in the  
 filamentous fungus Trichoderma reesei

JOURNAL J. Biol. Chem. 278 (34), 31988-31997 (2003)  
 MEDLINE 22803314

PUBMED 12788920

COMMENT

Genecor Intl.  
 925 Page Mill Road, Palo Alto, CA 94304, USA  
 Tel: (650) 846-7635  
 Fax: (650) 621-7817  
 Email: Pforeman@genecor.com  
 Seq primer: LT-F1 primer.

FEATURES

source

1..794  
 /organism="Hypocrea jecorina"  
 /mol\_type="mRNA"  
 /strain="QM6a"  
 /db\_xref="taxon:51453"  
 /clone="tric006xe08"  
 /dev\_stage="mycelia"  
 /clone\_lib="T.reesei mycelial culture, Version 3 april"  
 /note="Vector: PREP3Y; Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

Alignment Scores:

Pred. No.: 2,45e-20 Length: 794

Score: 338.00 Matches: 89

Percent Similarity: 51.97% Conservative: 43

Best Local Similarity: 35.04% Mismatches: 102

Query Match: 8.37% Indels: 20

DB: 6 Gaps: 9

US-09-917-376-3 (1-740) x CB896354 (1-794)

QY 422 ProMetValLys--GlyLeuGluThrAlaValAsnAspLeuSerProProSer 440

Db 58 CCACGGTCCGAGCGCATCGAGAAATTCCTCGTCCAGGACCTGGCTCTGCACCGCGC 117

QY 441 GlyAlaProLeuSerAlaLeuGlyAspLeuGlyGlyPheThrHisAla----- 457

Db 118 GGAAGCGAGCTATTGCGCGCAGTCGGAGACGACAAACGGTTCACCTTTGCCAGCAAAAC 177

QY 458 AspValThrAlaValProSerThrPheThrSerProValPheThrGlyThrSer 477

Db 178 GACCTCGGAGATCGCGCGCAGACGGTCTCGGCAACGCCACATCGGCGCACCTCGACGAGC 237



478 ValAspTyrAlaGluLeuAsnProSerIleValAlaGlySerPheAspProSer 497  
|||||  
238 GTGACTACCGCGGAACCTCGTCAAGAGCTGTCGCGTGGCGCAACACCGCGGACG 297  
|||||  
498 SerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGln 517  
|||||  
298 CAACAG-----GTGGCCATCTCGTCCGACGCGCGCGCGACGTGGAGCATC 342  
|||||  
518 GlySerGluProGlyGlyValThrGlyGlyThrValAlaAlaSerAlaAspGlySer 537  
343 GACTACGCGCGCGACGCTCATGAACGCGCGCGCGCGTGGCTATTCGCGCGCGCGAC 402  
538 ArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGly 557  
|||||  
403 ACGATCTCTGTCGACCGCTCGTCCGCG-----GTGGCGGCTCGCATGTTCCAG 453  
558 AsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgVal 577  
|||||  
454 GGCAGCTTGTCTCGTCTCGAGCTGCGCGCGCGCGCGTGCATCGCTCGACAGAAG 513  
578 AsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGly 597  
|||||  
514 ACCAACACGCTCTTACGCGCGCTCGGATCGACCTTTTACGTCAGCAAGACACCGCG 573  
598 ValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMet--- 616  
574 AGCAGCTTC-----ACGCGCGG-----CCAAAGCTGGCGCGCGCGAGGACGATCGG 621  
617 ---PheHisAlaValProGlyLysGlyAspLeuTrpLeuAlaAlaSerSerGlyLeu 635  
622 GATATCGTGTCTACCCGACACCGCGCGCGCGTGTGTATGTCGACCGCGTGGCAT 681  
636 TyrHisSerThrAsnGlyGlySerSerTrpSerAlaIle---ThrGlyValSerSerAla 654  
682 TTCCGCTCCACAGATCGCGCGCGACCTTTGGCCAAAGTCTCCACCGCGCTGACCAACACC 741  
655 ValAsnValGlyPheGlyLysSerAlaProGlySerSerTyr 668  
742 TACCAGATCGCTCGGCTGGTGGCTCA---GGCTCGAACTGG 780

RESULT 12  
BF072664 546 bp mRNA linear EST 18-OCT-2000  
LOCUS NCSM3H73 Subtracted Mycelial Neurospora crassa cDNA clone SM3H7 5'  
DEFINITION similar to avicelase III, Aspergillus aculeatus, mRNA sequence.  
ACCESSION BF072664  
VERSION BF072664.1 GI:10866169  
KEYWORDS EST.  
SOURCE Neurospora crassa  
ORGANISM Neurospora crassa  
REFERENCE  
AUTHORS Nelson, M.A., Kang, S., Braun, E.L., Crawford, M.E., Dolan, P.L.,  
Leonard, P.M., Mitchell, J., Armijo, A.M., Bean, L., Blueyes, S.,  
Cushing, T., Ertter, A., Fleharty, M., Gorman, M., Judson, K.,  
Miller, R., Ortega, J., Pavlova, I., Perea, J., Todisco, S.,  
Trujillo, R., Valentine, J., Wells, A., Werner-Washburne, M., Yazzie, S.  
and Natvig, D.O.  
Expressed sequences from conidial, mycelial, and sexual stages of  
Neurospora crassa  
JOURNAL Fungal Genet. Biol. 21, 348-363 (1997)  
MEDLINE 97435549  
PUBMED 9290248  
COMMENT Contact: Natvig, D.O./Nelson, M.A.  
Department of Biology  
University of New Mexico  
Castetter Hall, Albuquerque, NM 87131, USA  
Tel: 505 277 3411  
Fax: 505 277 0304  
Email: nsg@biology.unm.edu.  
Location/Qualifiers  
1. .546

FEATURES  
source

/organism="Neurospora crassa"  
/mol\_type="mRNA"  
/strain="74-OR23-IV A (FGSC 2489)"  
/db\_xref="taxon:5141"  
/clone="SM3H7"  
/sex="Mating type A"  
/tissue\_type="Mycelium"  
/dev\_stage="Mycelium"  
/lab\_host="E. coli"  
/clone\_lib="Subtracted Mycelial"  
/note="Vector: pBluescript SK (-); Site 1: EcoRI; Site 2:  
XhoI; 2% sucrose for 24 hours. cDNA directionally cloned  
into pBluescript SK(-) using the Uni-ZAP XR vector system  
(Stratagene, La Jolla, CA). previously identified highly  
expressed clones were subtracted from this library."

## ORIGIN

Alignment Scores:  
Pred. No.: 1.98e-18 Length: 546  
Score: 314.50 Matches: 72  
Percent Similarity: 54.26% Conservative: 30  
Best local Similarity: 38.30% Mismatches: 75  
Query Match: 7.79% Indels: 11  
DB: 2 Gaps: 3

US-09-917-376-3 (1-740) x BF072664 (1-546)

QY 358 SerAlaGluProTrpLeuThrPheGlyValGlnProAsnProValProSerProLys 377  
|||||  
Db 5 TCGGACGAGCTCGATCGAGACTGGCTTCCTTTTCAG-----GATACCAAGCAC 55  
QY 378 LeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyr 397  
|||||  
Db 56 CTCGCTGATGATCATCATCTTCGAGATCAACCCCTCGACGACGACCATTTGGCTCTAC 115  
QY 398 GlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGln 417  
|||||  
Db 116 GCCACGGTCTCACTGTTTACGGCGGCACGACCTGACCAAGTGGGACACCGTCCACAAC 175  
QY 418 IleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuSer 437  
|||||  
Db 176 GTGACATCCAGTCTTGGCGGTTGGCATCGAAGAAATGGCTGTCTCGCTCGCTCC 235  
QY 438 ProProSerGlyAlaProLeuLeuSerAlaLeuGlyAspLeuGlyGlyPheThr----- 455  
236 GCGCCCGGTGCTCGAGCTTCTCGCGCGCTGCGTGACGACTGCGGCTTCACCTTCAAG 295  
QY 456 ---HisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThr 474  
296 TCCAGCTCGACCTCGGACCTCTCCCAAGACACCTCGGATGACGCCCGGCGGCGCAGC 355  
QY 475 GlyThrSerValAspTyrAlaGluLeuAsnProSerIleValAlaGlySerPhe 494  
356 TCCGCGGAGCTCGACTACGCGCGCAACAAAGCCGCAACGCTGTCGTCATCGGCTCGGC 415  
QY 495 AspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyAsn 514  
416 TCCGCGGCGCGACAG-----GTGGCGGTATCTCTGGACGCGCGGCGCTCG 460  
QY 515 TrpPheGlnGlySerGluProGlyGlyValThrGlyGlyThrValAlaAlaSerAla 534  
461 TGGCATGTCTACAACGCGCGACCGACACCAAGACGAGCGGCGGCTCGCTCTCTGCC 520  
QY 535 AspGlySerArgPheValTrpAla 542  
521 GAGCGCGACACCATCGTCTGGTCT 544

RESULT 13  
BU639045  
LOCUS mgcw011xD18f.b RCW Lambda Zap Express Library Magnaporthe grisea  
DEFINITION cDNA clone mgcw011xD18 5', mRNA sequence.  
ACCESSION BU639045



```

VERSION      BU639045.2  GI:30397224
KEYWORDS     EST
SOURCE       Magnaporthe grisea (anamorph: Pyricularia grisea)
ORGANISM     Magnaporthe grisea
REFERENCE    Rukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
AUTHORS      Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
TITLE        1 (bases 1 to 389)
              Ebbole,D.J., Yuan,J., Thomas,T.L., Bobrowicz,P., Lu,G.,
              Bhattraai,K. and Dean,R.A.
              Expressed sequence tags from the rice blast fungus, Magnaporthe
              grisea
JOURNAL      Unpublished (2002)
COMMENT      On Sep 30, 2002 this sequence version replaced gi:23351371.
              Contact: Ebbole DJ
              Department of Plant Pathology & Microbiology
              Texas A&M University
              Peterson Bldg, MS2132, College Station, TX 77843-2132, USA
              Tel: 979 845 4831
              Fax: 979 845 6483
              Email: d-ebbole@tamu.edu
              Chromatogram file of this sequence is available, see contact
              person; Best nr hit (April. 22, 2003) gb|EAA29333.1| hypothetical
              protein [Neurospora crassa] 206 4e-53
              PCR Primers
              FORWARD: T3 primer
              BACKWARD: T7 primer
              Plate: mgcw011 row: D column: 18
              Seq primer: T3.

FEATURES
    source
        1..389
            /location/Qualifiers
                /organism="Magnaporthe grisea"
                /mol_type="mRNA"
                /strain="CP987"
                /db_xref="taxon:148305"
                /clone="mgcw011xD18"
                /sex="Mat1-1 hermaphrodite"
                /tissue_type="Mycelium"
                /dev_stage="Day 5 post-inoculation"
                /clone_lib="RCW Lambda Zap Express Library"
                /note="Vector: phluescript excised from Lambda Zap
                Express; Site 1: EcoRI; Site 2: XhoI; Day 5
                post-inoculation MRNAs prepared from Magnaporthe grisea
                grown at 23C in the dark with constant gyratory shaking
                100 rpm in Vogel's minimal medium containing 0.5% isolated
                rice cell walls as the sole carbon source. Library
                provided by Sheng-Cheng Wu. Sequences were processed by
                one of two methods. Where a full-length alignment to the
                M. grisea genome sequence was available, the EST sequence
                was trimmed according to the alignment, otherwise sequence
                quality was assessed using phredPhrap version 991019 and
                trimmed according to phd files (0.05) and for vector
                seqs."

ORIGIN
Alignment Scores:
Pred. No.:      6.41e-18      Length:      389
Score:          306.50      Matches:      55
Percent Similarity: 66.67%      Conservative: 27
Best Local Similarity: 44.72%      Mismatches: 38
Query Match:    7.59%      Indels:      3
DB:             5          Gaps:      1

US-09-917-376-3 (1-740) x BU639045 (1-389)

Qy 334 GlyAlaThrTpThrArgIleTtrAspThrThrSerTyrProAsnArgSerLeuArgTyr 353
Db 5 GGGGCTACATGTCGAGATCTTGGAGTGGACGAGTATCCGACATGACGACTACTAC 64
Qy 354 ValLeuAspIleSerAlaGluProTrrPheGlyValGlnProAsnProProVal 373
Db 65 TCGATCAACACAGACAGCGCCCTGGATAGTCTGGC-----CATCTTTCGAGG 115
Qy 374 ProSerProLysLeuGlyTrrPheMetAspGluAlaMetAlaIleAspProPheAsnSerAsp 393

```

---

```

Db 116 GATTCGAAGCGCCTTGGTTGGATGATCGAGGCCCTTGAATCGACCTGCGATCGTAC 175
Qy 394 ArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrrAsp 413
Db 176 CATTTGGCTTTACGGAACTGGGCTCTCGATCTTTGGGGGACATGACTTGAACCAATGGGAC 235
Qy 414 SerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsn 433
Db 236 ACGGTTCAATACATATCGATTCACTCTTTGGCCGATGGTGTGGAAGAACTGCAGTTCG 295
Qy 434 AspLeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGly 453
Db 296 GAGATGGCTTCAGCGCCCGAGGCTCCGAACTACTTGGCGCAATGGGACATTACTGGA 355
Qy 454 PheThrHis 456
Db 356 TTCATATAC 364

RESULT 14
CF875815 713 bp mRNA linear EST 31-OCT-2003
LOCUS     trico39xxk06.b1 T.reesei mycelial culture, Version 6 October 2003
DEFINITION Hypocrea jecorina cDNA clone trico39xxk06, mRNA sequence.
ACCESSION CF875815
VERSION   CF875815.1 GI:38130497
KEYWORDS  EST.
SOURCE    Hypocrea jecorina (anamorph: Trichoderma reesei)
ORGANISM  Hypocrea jecorina
REFERENCE 1 (bases 1 to 713)
AUTHORS   Diener,S.E., Dunn-Coleman,N., Foreman,P., Houfek,T.D.,
            Teunissen,P.J.M., van Solingen,P., Dankmeyer,L., Mitchell,T.K.,
            Ward,M. and Dean,R.A.
TITLE     Characterization of the protein processing and secretion pathways
            in a comprehensive set of expressed sequence tags from Trichoderma
            reesei
JOURNAL   FEMS Microbiol. Lett. 230 (2), 275-282 (2004)
COMMENT   Contact: Ralph A. Dean
            Fungal Genomics Laboratory
            North Carolina State University
            Campus Box 7251, Raleigh, NC 27695, USA
            Tel: 919-513-0020
            Fax: 919-513-0024
            Email: ralph.dean@ncsu.edu
            Seq primer: LT-F1 primer.

FEATURES
    source
        1..713
            /organism="Hypocrea jecorina"
            /mol_type="mRNA"
            /strain="QM6a"
            /db_xref="taxon:51453"
            /clone="trico39xxk06"
            /dev_stage="mycelia"
            /clone_lib="T.reesei mycelial culture, Version 6 October
            2003"
            /note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial
            culture grown from 24 hrs to 6 days with varying Carbon
            and Nitrogen sources and concentrations."

ORIGIN
Alignment Scores:
Pred. No.:      1.11e-15      Length:      713
Score:          286.50      Matches:      81
Percent Similarity: 48.79%      Conservative: 40
Best Local Similarity: 32.66%      Mismatches: 106
Query Match:    7.10%      Indels:      21
DB:             7          Gaps:      10

US-09-917-376-3 (1-740) x CF875815 (1-713)

Qy 458 AspValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSer 477

```



Db	10	GACCTCGGTACATCGCCGCGAGTGTGCTGGGCAACGCCACATCGGGCCACCTTCGACGAGC	69
Qy	478	ValAspTyrAlaGluLeuAenProSerIleIleValArgAlaGlySerPheAspProSer	497
Db	70	GTGAGTACCGCGGAACTCGGTCAAGAGCGTGTGTCGCGTGGGCAACACCGCCGCGACG	129
Qy	498	SerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGln	517
Db	130	CAACAG-----GTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATC	174
Qy	518	GlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySer	537
Db	175	GACTACGCGCGCACACGTCCATGAACGGCGCACGCTGGGCTATTGCGCCGACGCGCAC	234
Qy	538	ArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGly	557
Db	235	ACGATCTCTGGTCCGACCGCCTCGTCCGCG-----GTGCACGCTCGCAGTTC	285
Qy	558	AsnSerTrpAlaAlaSerGlnGlyValProAlaAlaGlnIleArgSerAspArgVal	577
Db	286	GGCAGCTTGTGCTCCGTCTCGAGCGCTGCGCGGCGCGCTCATCGCTCGACACGAAG	345
Qy	578	AsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGly	597
Db	346	ACCAACAGCGTCTTCTACGCGCGCTCCGATCGACCTTTTACGTGACGACGACACCGCG	405
Qy	598	ValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyValAlaValGlyValMet	616
Db	406	AGCAGCTTC-----ACGCGCGG---CCCAAGCTGGGCGAGCGAGGACGATCCGG	453
Qy	617	---PheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeu	635
Db	454	GATATCGCTGCTCACCCGACCAACCGCGGCGACGTTGTATGTCGACCGACGTCGCGCAT	513
Qy	636	TyrHisSerThrAsnGlyGlySerSerTrpSerAlaIle---ThrGlyValSerSerAla	654
Db	514	TTTCGCTCCACAGACTCGGGGACGACCTTTGGCCCAAGTCTCCACCGCCCTGACCAACCC	573
Qy	655	ValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValVal	674
Db	574	TACAGATCGCCCTGGGTGGGCTCA---GGCTCGAACTGG---AACCCTGATGCTTC	627
Qy	675	GlyThrIleGlyGlyValThrGlyAla-----TyrArgSerAspAspCysGlyThrThr	692
Db	628	GGCACC-----GGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGCGCTCC	681
Qy	693	TrpValLeuIleAsnAspAspGln	700
Db	682	TGGACGGACATTACGGGCTCCACG	705

RESULT 15	CB904767	782 bp	mrna	linear	EST 02-JUL-2003
LOCUS	CB904767				
DEFINITION	tric039xx06 T.reesei mycellial culture, Version 3 april Hypocrea				
	jecorina cDNA clone tric039xx06, mRNA sequence.				
ACCESSION	CB904767				
VERSION	CB904767.1	GI:30119425			
KEYWORDS	EST.				
SOURCE	Hypocrea jecorina (anamorph: Trichoderma reesei)				
ORGANISM	Hypocrea jecorina				
	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;				
	Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocreae.				
REFERENCE	1 (bases 1 to 782)				
AUTHORS	Foreman,P.K., Brown,D.B., Dankmeyer,L., Dean,R., Diener,S., Dunn-Coleman,N.S., Goedegebuur,F., Houfek,T.D., England,G.J., Kelley,A.S., Meerman,H.J., Mitchell,T., Mitchinson,C., Olivares,H.A., Teunissen,P.J., Yao,J. and Ward,M.				
TITLE	Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus Trichoderma reesei				
JOURNAL	J. Biol. Chem. 278 (34), 31988-31997 (2003)				
MEDLINE	22803314				
PUBMED	12788920				

**COMMENT**  
 Contact: Pamela K. Foreman  
 Genencor Intl.  
 925 Page Mill Road, Palo Alto, CA 94304, USA  
 Tel: (650) 846-7635  
 Fax: (650) 621-7817  
 Email: pforeman@genencor.com  
 Seq primer: LT-P1 primer:  
 Location/Qualifiers  
 source  
 1..782  
 /organism="Hypocrea jecorina"  
 /mol\_type="mRNA"  
 /strain="QM6a"  
 /db\_xref="taxon:51453"  
 /clone\_lib="trio39xx06"  
 /dev\_stage="mycelia"  
 /clone\_lib="T. reesei mycelial culture, Version 3 april"  
 /notes="Vector: pREP3Y; Site 1: Not 1/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

**FEATURES**  
 source  
 1..782

**ORIGIN**

Alignment Scores:					
Pred. No.:	1..28e-15	Length:	782		
Score:	286.50	Matches:	81		
Percent Similarity:	48.79%	Conservative:	40		
Best Local Similarity:	32.66%	Mismatches:	106		
Query Match:	7.10%	Indels:	21		
DB:	6	Gaps:	10		

US-09-917-376-3 (1-740) x CB904767 (1-782)

QY	458	AspValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSer	477
DB	79	GACCTCGTACATCGCGCAGTAGTGGTCTGGCAACGCCACATGGGCCACTCGACGAGC	138
QY	478	ValAspTyrAlaGluLeuAsnProSerIleValArgAlaGlySerPheAspProSer	497
DB	139	GTCGACTACGCGGGAACCTCGGTCAAGAGCGTCTGTCGCGTTCGGCAACACCGCCGGCAGC	198
QY	498	SerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGln	517
DB	199	CAACAG-----GTGGCCCATCTCGTCCGCGCGCGCGCGCGAGCTGGAGGATC	243
QY	518	GlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySer	537
DB	244	GACTACGGCGCCGACAGCTCATCAACGGCGGCAAGGTGGCCATTTCGGCCGAGCGGCAC	303
QY	538	ArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGly	557
DB	304	ACGATCCTCTGGTCGACCGCTCGTCCGGC-----GTGACGCGCTGCGAGTTCCAG	354
QY	558	AsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgVal	577
DB	355	GGCAGCTTGGCTTCGGTCTCGAGCTGCCCGCGGCGCGGTCATCGCTCGGACCAAGAG	414
QY	578	AsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGly	597
DB	415	ACCAACAGCGTCTTCTACGCGCGGCTCCGGATCGACCTTTTACGTACCAAGGACACCGGC	474
QY	598	ValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMet---	616
DB	475	AGCAGCTTC-----ACGCGCGGG--CCCAAGCTGGCGAGCGGAGGACGATCCGG	522
QY	617	---PheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeu	635
DB	523	GATATCGTGTCTACCCCGACCCACCGCGGCGCAGTTGTATGTCTTCGACCGGACGTCGGGATA	582
QY	636	TyrHisSerThrAndGlyGlySerSerTrpSerAlaIle---ThrGlyValSerSerAla	654
DB	583	TTCCGCTCCACAGACTCGGGGACGACCTTGGCCCAAGTCTCCACCGGCTCACCACACACC	642
QY	655	ValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheVal	674



```

Db      643 TACCAGATCGCCCTGGGTGTGGCTCA---GGCTCGAACTGG---AACCTGTATGCCTTC 696
Qy      675 GlyThrIleGlyGlyValThrGlyAla-----TyrArgSerAspAspCysGlyThrThr 692
Db      697 GGCACC-----GGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCC 750
Qy      693 TrpValLeuIleAsnAspAspGln 700
Db      751 TGGACGGACATTCAGGGCTCCAG 774

```

Search completed: October 5, 2005, 21:09:27  
 Job time : 7126.69 secs







Qy 39 GlyGlyMetTyrArgTrpAspAlaalaasnGlyArgTrpIleProLeuLeuAspTrpVal 58  
Db 2933 GGC-----GGCAACCGGATAGGCGAGGTCAACCGCGGAC 2901  
Qy 59 GlyTrpAsnAsnTrpGlyTyr-----AsnGly-----ValVal 69  
Db 2900 GGCAGCGCAACTGGAGCTTCACGCGCGGACGCGCGCGGCAACGCGCAGGTGATCGTC 2841  
Qy 70 SerIleAlaAlaAspProIle---AsnThrAsnLysValTrpAla-----Ala 84  
Db 2840 GCCAGCGCCACCGACCGGCAATACCGGCGCGGCGGCGGCGCACCGTGGAGCGG 2781  
Qy 85 ValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu----- 100  
Db 2780 GTGGCGCGCGCGCGGTGATCGATCCGAGCAACGCGCACCATCAGCGGACCGCG 2721  
Qy 101 -----ArgSer 102  
Db 2720 GAGCGCGGCGCAAGTGTCTCTACCGACGCGCAACCGGATCGGCGAACCAC 2661  
Qy 103 SerAspGlnGlyAlaThrTrpGlnIleThrPro---LeuProPheLysLeuGly----- 119  
Db 2660 GCCAGCGGACGGGCAACTGACCTTCACGCGCGCGGCGGCGGCGGCAACGCGCAGGTG 2601  
Qy 120 -----GlyAsnMetProGlyArgGlyMetGlyGluArg 130  
Db 2600 GTCAACCGCGTGGCCAGGACCTCGGCGCAATACCGGCGCGGCGGCGGCGGCGGCGG 2541  
Qy 131 LeuAlaValAspProAsn-----AsnAspAsnIleLeuTyrPhe 143  
Db 2540 GACCGGTGGCGCGGCAACAGCCTGTGTCAATCCGAGCAACGCGCAACTGCTCAACGGT 2481  
Qy 144 GlyAlaProSerGlyLysGlyLeu-----Trp 152  
Db 2480 ACCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2421  
Qy 153 ArgSerThrAspSerGlyAlaThrTrp-----SerGlnMetThrAsnPhe 167  
Db 2420 ACCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2364  
Qy 168 ProAspValGlyThrTyrIle---AlaAsnProThrAspThrThrGlyTyrGlnSerAsp 186  
Db 2363 -----GGCAGCGGTGTCAACGTCACCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGG 2316  
Qy 187 IleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGln----- 204  
Db 2315 -----GCTCCCGTACCCAGCGGTGGATTCTCGTGGCGGCGGCGGCGGCGGCGGCGG 2262  
Qy 205 ---AlaSerLysThrIlePheValGlyValAlaAlaAspProAsnAsnProValPhe----- 221  
Db 2261 CCGAGCAACGGTTCGGTGATCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2202  
Qy 222 -----TrpSerArgAspGlyGlyAlaThrTrpGlnAla 232  
Db 2201 GATGCAACGCAACCCGATTGGCGAGTCCACCGCGGAGCGGCGGCGGCGGCGGCGGCGG 2142  
Qy 233 ValProGlyAlaPro----- 237  
Db 2141 ACTCCAGCATCCCGCTGCGGATGGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2082  
Qy 238 -----ThrGlyPheIleProHisLysGlyVal 246  
Db 2081 AATGTGACAGTGGCGCGGTGATCACTGTGATGGCGGTGGCGGCGGCGGCGGCGGCGGCGG 2022  
Qy 247 PheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrThrGlyProTyr--- 265  
Db 2021 ATCGATCCGACACCGGACCGGAGTAAAGCGGTACCGGCGGAGCGGCGGCGGCGGCGGCGG 1962  
Qy 266 -----AspGlySerSerGlyAspValTrpLysPheSerValThr---SerGlyThrTrp 282  
Db 1961 CTCACCGATGGCGGCGGCAACCCGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1902  
Qy 283 ThrArgIleSerProValPro-----AlaAsnAlaGlnIleArgSer----- 574

Db 1901 ACGTTCACCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1842  
Qy 293 ThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisPro----- 310  
Db 1841 CCGCGCGGCAATACACGCGGTCCGCGGAGGTTCACCGTCGATGCCATCCGCCCGCGGCGG 1782  
Qy 311 -----AsnThrIleMetValAlaThrGlnIleSerTrpTrpProAsp 324  
Db 1781 CCGGTGATCAATCCGAGCAACGGAGTCGTATCAGCGGTACGCGGGAACCGGCGGCGGCGG 1722  
Qy 325 ThrIleIlePheArgSerThrAspGlyGly-----AlaThrTrpThr 338  
Db 1721 GTGATCTCT-----ACCGAGCGCAACCGCAACCGGATCGGCGGAGTCACCGCGGAC 1671  
Qy 339 ArgIleTrpAspTrpThr-SerTyrProAsnArg-----SerIle 351  
Db 1670 GGCAGCGCAAGTGGGTTCACGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1611  
Qy 351 uArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnPr 371  
Db 1610 CGGTGG-----CCGAGGACCGCGCGGCAACACAGCAGTC-----CC 1572  
Qy 371 oProValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAs 391  
Db 1571 ACCAGCGCA---CGTCACTCGGTGGCGGCGGCAACGCGGCGGCGGCGGCGGCGGCGG 1516  
Qy 391 nSerAspArgMetLeuTyrGlyThr-----GlyAlaThrLeuTyrAlaThrAsnAs 408  
Db 1515 CGGTAGC---GTGATCGCGGTACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1464  
Qy 408 pLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuG 428  
Db 1463 -----GACGGCAACGGGCAAC-----CCGATCGGCGGCGGCGGCGGCGGCGG 1429  
Qy 428 uGluThrAlaValAsnAspLeuIleSerProSerGlyAlaProLeu----- 444  
Db 1428 CGATGGCAGCGGCAACTGGAGCTTCACGCGC-----GGCAGCGCGCTGCTCAATGGCAC 1375  
Qy 445 ---IleSerAlaLeuGly---AspLeuGlyGlyPheThrHisAlaAspVal---ThrAl 461  
Db 1374 GGTGTCATCGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1315  
Qy 461 aValProSerThrIlePheThrSerProValPhe-----ThrThrGlyThrSerVal-- 478  
Db 1314 GGTGACTCGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1255  
Qy 479 -AspTyrAlaGluLeuAsnProSerIleValArgAlaGlySerPheAspProSerSe 498  
Db 1254 CGGTACCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1195  
Qy 498 rGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrp----- 515  
Db 1194 CCAGGCGC-----ACCGCGGATGGCAGCGGCACTGGAGCTTCAC 1156  
Qy 516 -----PheGlnGlySerGluProGly 523  
Db 1155 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1096  
Qy 523 yValThrGlyGlyThr----- 529  
Db 1095 CAATACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1036  
Qy 530 -----ValAlaAlaSerAlaAspGlySerArgPheValTrpAl 542  
Db 1035 CAACCGGAGCAACGCGCAGCGGTGATCGCGGTACCGGGAACCGGCGGCGGCGGCGGCGG 976  
Qy 542 aProGlyAspProGlyGlnProVal-----ValTyrAlaValGlyPheGlyAsnSerTr 560  
Db 975 CACCGAGCGGCGGCGGCAACCCGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 919  
Qy 560 pAlaAlaSerGlnGlyValPro---AlaAsnAlaGlnIleArgSer----- 574



Db 918 GAGCTTCACGGCCGCGCAGCGCGTGGCAACAGCGCTCGTGATCAATCGCTGGGCCAGGA 859  
 Qy 575 -----AspArgValAsnProLysThr 581  
 Db 858 CGCGCGCGCAACACACAGCGCGCGCGCAGCACCGTGGACTCGGTAGCCCGGCCAC 799  
 Qy 581 r---PheTyrAlaLeuSerAsnGlyThrPheTyrArgSer----- 593  
 Db 798 CCGGTGCTGATCGAGCAACGGTACGTGATCAGCGGTACCGCGAAGCGGGGCCAC 739  
 Qy 594 -----ThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeu-ProSers 610  
 Db 738 GGTGATCTCTACCGCGCGCGC-----GGCAACCGATACCGCAGCGCCCGATGG 685  
 Qy 610 erglyAlaValGlyValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuA 630  
 Db 684 CAGCGGCAACTGGAGCTTCACTCCGGGCACA-CGGCTGACCAACGCGCAGCGTGTCAATG 626  
 Qy 630 laAlaSerSerGlyLeuTyrHisSerThrAsnGly-----GlySerSerTrpSerAlaI 648  
 Db 625 CGGTGGCCAGGACCGCGCGCAACACCGCGGTCCGGTCAGCACCACACATGGAGCGCGG 566  
 Qy 648 leThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSert 668  
 Db 565 TGGCCCCCGCCACCCCGGTGATCGACCGCAATGGTGTCAAACTCAGCGCGCACCGCGG 506  
 Qy 668 yrProAlaValPheValGlyThrIleGlyGly----- 679  
 Db 505 AACCGCGGTCCGGGTGATCTCCAGTGGCAATGGCAACCGATCGCCAGACCTTCG 446  
 Qy 680 -----ValT 681  
 Db 445 CCGAGGTAGCGGCAACTGGACCTTCACACCGGCGCACCGCGTGGCCACGCGGTGG 386  
 Qy 681 hrGlyAlaTyrArgSerAspAspCysGlyThrTrpValLeuIleAsnAspAspGlnH 701  
 Db 385 TCACCGCGTGGCCCGCAGCGCGCGCGCAATACAG-----CGGTCCGCGCAGC 336  
 Qy 701 isGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArg-Val 720  
 Db 335 ACCAGGTGGATACGGTGGCGCGCGCCACCGCGGTGATCAATGCCAGCAACGCGACGTG 276  
 Qy 721 TyrIleGlyThrAsn-----GlyArgGlyIleValTyrGlyAspIleGlyAlaPro 738  
 Db 275 ATCACCAGCGCGCGAGTGGCGCGCAAGTGTCTCTCAGCGCGCAACGCGCAACCGG 216  
 Qy 739 SerGly 740  
 Db 215 ATCGGC 210

RESULT 2  
 US-09-252-991A-13656  
 ; Sequence 13656, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 13656  
 ; LENGTH: 8211  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 ; US-09-252-991A-13656  
 Alignment Scores:

Pred. No.: 2,98e-10 Length: 8211  
 Score: 227.50 Matches: 218  
 Percent Similarity: 31.30% Conservative: 90  
 Best Local Similarity: 22.15% Mismatches: 360  
 Query Match: 5.64% Indels: 319  
 DB: 4 Gaps: 52  
 US-09-917-376-3 (1-740) x US-09-252-991A-13656 (1-8211)  
 Qy 20 AspGlyIleValPheAsnGluGlyAla---ProGlyIleLeuTyr-ValArgThrAspIle 38  
 Db 3574 AACCGCTCGTCATCAGCGGCACCGCCGAGCGCGTGCACCGTACCTTACCGATGCC 3633  
 Qy 39 GlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpVal 58  
 Db 3634 GGC-----GGCAACCGATAGGCGAGGTCAACCGCGCAG 3666  
 Qy 59 GlyTrpAsnAsnTrpGlyTyr-----AsnGly-----ValVal 69  
 Db 3667 GGCAGCGCAACTGGAGCTTCACGCGCGGCACCGCGCGCCACCGGCACCGGTGATCGTC 3726  
 Qy 70 SerIleAlaAlaAppProIle---AsnThrAsnLysValTrpAla---Ala 84  
 Db 3727 GCCACGCGCACCGACCCGACCGCAATACCGCGCCGCGAGCGCCACCAACCGGTGACGCG 3786  
 Qy 85 ValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu----- 100  
 Db 3787 GTGGCGCGCGCGCGCGTGTGATCGATCCGAGCAACGCGCACCATCAGCGCGCACCGCG 3846  
 Qy 101 -----ArgSer 102  
 Db 3847 GAGCGCGCGCCAGGTGATCTCACCAGCGCAACCGCAACCCGATCGCGCAACCAACC 3906  
 Qy 103 SerAspGlnGlyAlaThrTrpGlnIleThrPro---LeuProPheLysLeuGly----- 119  
 Db 3907 GCCACGCGCAGCGCGCAACTGGACCTTCACGCGCGCACCGCGCTGGCCCAACCGCGCGT 3966  
 Qy 120 -----GlyAsnMetProGlyArgGlyMetGlyGluArg 130  
 Db 3967 GTCAACGCGGTGGCCAGACCCCTGGGGCAATACCGGCGCGCGGCGGCGACTACCGTG 4026  
 Qy 131 LeuAlaValAspProAsn-----AsnAspAsnIleLeuTyrPhe 143  
 Db 4027 GACGCGGTGGCGCGCAACACGCGCTGTGGTCAATCCGAGCAACGGCAACCTGTCTCAACGGT 4086  
 Qy 144 GlyAlaProSerGlyLysGlyLeu-----Trp 152  
 Db 4087 ACCGCGCGAGCGCGCGCAGCACCGTGTGACCGAGCGCAACCGCAACCGATCGGCCAG 4146  
 Qy 153 ArgSerThrAspSerGlyAlaThrTrp-----SerGlnMetThrAsnPhe 167  
 Db 4147 ACCACCGCGCATGGCAGCGCGCAACTGGAGCTTCACGCGCGGTCTCGCACTACCCCAAC--- 4203  
 Qy 168 ProAspValGlyThrTyrIle---AlaAsnProThrAspThrThrGlyTyrGlnSerAsp 186  
 Db 4204 -----GGCACCGTGTCAACGTGACGCGGAGCGAGCGCGCGCGGCAATACCAAGC--- 4251  
 Qy 187 IleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGln----- 204  
 Db 4252 -----GCTCCCGCTACCAACGCGGTGATTCCTCGCTCGCTCGATCCCGCAGGTGGAT 4305  
 Qy 205 ---AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPhe----- 221  
 Db 4306 CCGAGCAACCGTTCCGTGTGATCAGCGCGCACCGCGCGCGCGGCAACCACTCATCATCAACC 4365  
 Qy 222 -----TrpSerArgAspGlyGlyAlaThrTrpGlnAla 232  
 Db 4366 GATGGCAACGCGCAACCCGATTTGGCCAGGTCAACCGCGCGAGCGCGGTAACTGCTCTTC 4425  
 Qy 233 ValProGlyAlaPro----- 237  
 Db 4426 ACTCCAGGATCCCGTCCCGGATGGCACGGTGGTCAACGTGGTGGCGCGCAGCCCAAGC 4485



QY 238 -----ThrGlyPheIleProHisLysGlyVal 246  
Db 4486 AATGTCGACAGTGGCGCGGGTGTACTGTGATGGCGTGGCCCGCGCGCGCGTGTG 4545  
QY 247 PheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyr--- 265  
Db 4546 ATCGATCCGAGCAACGCGCAGAGTAAGCGTACCGCGAGGCGCGCGCGTGTATC 4605  
QY 266 -----AspGlySerSerGlyAspValTyrPheSerValThr---SerGlyThrTrp 282  
Db 4606 CTCACCGATGGCGCGGCAACCCGATCGGCCAGCCACCGCGCGAGCGGCAACTGG 4665  
QY 283 ThrArgIleSerProValPro-----SerThrAsp 292  
Db 4666 ACGTTACCCCGGCGCACCCCGTGGCCNACGGCACCGGTGATCAACGCCGTGGCCGAGAC 4725  
QY 293 ThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisPro----- 310  
Db 4726 CCGGCGGCAATACCGAGCGTCCGCGCAGCGTACCGTGCATGCCATCGCCCGCGCGG 4785  
QY 311 -----AsnThrIleMetValAlaThrGlnIleSerTrpTrpProAsp 324  
Db 4786 CCGGTGATCAATCCGAGCAACGGAGTGTCTCATCGCGGTACGGCGGAAGCCGGGCCACG 4845  
QY 325 ThrIleIlePheArgSerThrAspGlyGly-----AlaThrTrpThr 338  
Db 4846 GTGATCTTC-----ACCGAGCGNACGGCAACCGCATCGGCCAGGTCAACGCCGAC 4896  
QY 339 ArgIleTrpAspTrpThr-SerTyrProAsnArg-----SerLeu 351  
Db 4897 GGCAGCGCAAGTGGGCTTTCAGCGCGCCACGCGTGGCCAAATGGCACGGTGTATCAAT 4956  
QY 351 uArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnPr 371  
Db 4957 CGCGTGG-----CCGAGAGCGCCCGCGCAACACAGCAGTCT-----CC 4995  
QY 371 oProValProSerProLysLeuGlyTyrMetAspGluAlaMetAlaIleAspProPheAs 391  
Db 4996 ACCAGCGCA-----CGTGCATCGGTGGCGCAGCA-GCCCGGTGATCGATCCGAGCA 5051  
QY 391 nSerAspArgMetLeuTyrGlyThr-----GlyAlaThrLeuTyrAlaThrAsnAs 408  
Db 5052 CGGTAGC---GTGATCGCGGTACCGCGAGCGTGTGTGCCAGCGGTGATCTCAACC- 5103  
QY 408 pLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuG 428  
Db 5104 -----GACGGCAACGGCAAC-----CCGATCGCGCAGGTCAACCGC 5138  
QY 428 uGluThrAlaValAsnAspLeuIleSerProSerGlyAlaProLeu----- 444  
Db 5139 CGATGGCAGCGCAACTGGAGTTCACGCC-----GGCAGCGCGTGTCCAATGGCAC 5192  
QY 445 -----IleSerAlaLeuGly---AspLeuGlyGlyPheThrHisAlaAspVal---ThrAl 461  
Db 5193 GGTGTCTCAATGGGTGGCCAGGAGCGTGGCGGCAACACCGCGCCCGCGCGCAGCACAC 5252  
QY 461 aValProSerThrIlePheThrSerProValPhe-----ThrThrGlyThrSerVal-- 478  
Db 5253 GGTGACTCGGTGGCGCGCGCGCGTGTATCGACCCGAGCAACGGCGGTGATCGC 5312  
QY 479 -AspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSe 498  
Db 5313 CGGTACCGCGAAGCCGTGGCAGCGGTGATCTCATCCGATGGCGCGCAACCCGATCGG 5372  
QY 498 rGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrp----- 515  
Db 5373 CCAGGCC-----ACCGCGATGGCAGCGCAACTGGAGTTCAC 5411  
QY 516 -----PheGlnGlySerGluProGlyG 523  
Db 5412 CCGCGGCGCGCGTGGCCAAACGGCACCGGTGTATCAATCGGTGGCGCCAGGATCCGCGCG 5471  
QY 523 yValThrThrGlyGlyThr----- 529

Db 5472 CAATACGAGCGCGCGCGACGACCAACCGGTGGACCGCGTGGCCCGCGCCACCCCGGTGT 5531  
QY 530 -----ValAlaAlaSerAlaAspGlySerArgPheValTrpAl 542  
Db 5532 CAACCCGAGCAACCGCAGCGTGTACCGCGGAGCGCGCGCGCGCGTATCTCT 5591  
QY 542 aProGlyAspProGlyGlnProVal-----ValTyrAlaValGlyPheGlyAsnSerTr 560  
Db 5592 CACCGAGCGCGCGCAACCCGATCGGCCAGGTTCACCGCGCGCGCGCGCGCAAC---TG 5648  
QY 560 pAlaAlaSerGlnGlyValPro---AlaAsnAlaGlnIleArgSer----- 574  
Db 5649 GAGTTCACGCGCGCGCACCGCGTGGCCAAACCGTTCGGTGTCAATGGCGTGGCGCCACG 5708  
QY 575 -----AspArgValAsnProLysThr 581  
Db 5709 CGCCCGCGCGCAACACCGCGCGCGCGCGCACCGAGTGTGATCGCGTACCGCGGAGCGCGCCAC 5768  
QY 581 r---PheTyrAlaLeuSerAsnGlyThrPheTyrArgSer----- 593  
Db 5769 CCGGTGTCTCGATCCGAGCAACCGGTACGTGTATCGAGCGTACCGCGGAGCGCGCGCGCCAC 5828  
QY 594 -----ThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeu-ProSerS 610  
Db 5829 GGTGATCTCTACCGAGCGCGCGC-----GGCAACCCGATACCGCGAGCGCGCGCGATGG 5882  
QY 610 erGlyAlaValGlyValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuA 630  
Db 5883 CAGCGCGCAACTGGAGCTTCTACCTCCGCGGCACA-CCGCTGACCAACCGCGCGGTGATCAATG 5941  
QY 630 laAlaSerSerGlyLeuTyrHisSerThrAsnGly-----GlySerSerTrpSerAlaI 648  
Db 5942 CGTGGCCCGAGCAGCGCCCGCGCAACACCGCGGTCCCGTCCAGCACACACGCGCGCG 6001  
QY 648 leThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerT 668  
Db 6002 TGGCCCGCGCGCACCCCGGTGTATCGACCGAGCAATGGTGTCAAACTCAGCGCGCACCGCG 6061  
QY 668 yrProAlaValPheValValGlyThrIleGlyGly----- 679  
Db 6062 AACCGCGGTCTCGGTGTATCTCTACCGATGGCAATGGCAACCCGATCGGCCAGACCTCG 6121  
QY 680 -----ValT 681  
Db 6122 CCGAGGTAGCGGCAACTGGACCTTCACACCGGGCACCGCGTGGCCCAACGGCACCGTGG 6181  
QY 681 hrGlyAlaTyrArgSerAspAspCysGlyThrTrpValLeuIleAsnAspAspGlnH 701  
Db 6182 TCNACCGCGTGGCCCGCGAGCACCGCGCGCGCAATACCG-----CGTCCGCGCGCAGC 6231  
QY 701 isGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArg-Val 720  
Db 6232 ACCAGGTGGATACCGTGGCGCGCGCGCGCGCGTGTATCAATGCCAGCAACCGCGCGGTG 6291  
QY 721 TyrIleGlyThrAsn-----GlyArgGlyIleValTyrGlyAspIleGlyAlaPro 738  
Db 6292 ATCACCAGCGACCGCGAGGTGGCGCGCAAGTGTCTTCCCGAGCGCAACCGCAACCCG 6351  
QY 739 SerGly 740  
Db 6352 ATCGGC 6357

## RESULT 3

US-09-252-991A-13774/c  
; Sequence 13774, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A



```

CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 13774
LENGTH: 4188
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13774

```

Alignment Scores:	
Pred. No.:	2.05e-08
Score:	203.00
Percent Similarity:	32.38%
Best Local Similarity:	22.22%
Query Match:	5.03%
DB:	4
Gaps:	46
Indels:	274
Mismatches:	297
Conservative:	85
Matches:	186
Length:	4188

US-09-917-376-3 (1-740) x US-09-252-991A-13774 (1-4188)

Qy	47	AlaAsnGlyArgTrrp	-----IleProLeuLeuAspTrpValGlyTrpAsn	61
Db	2496	GCCACGGCAACTGGTCTTTACGCGCTGCACCGCTGCGGAGC--GTACCGTGGTCA	2439	
Qy	62	AsnTrp-----GlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsn	79	
Db	2438	ACGTGTGGCCAGGGATGCGCGGCACACGACAGTCCGCGCCACAGGTTA-----	2388	
Qy	80	LysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIle	99	
Db	2387	---CCGTGGAT--GCCGTGCGCGCCGACCCACCGCTCGATCCGAGCAACGGTACGACC	2332	
Qy	100	LeuArg---SerSerAspGlnGlyAlaThaTrpGlnIleThrProLeuProPheLysLeu	118	
Db	2331	CTCAGCGGCACCGCCGAGCGCGGCTACCGTGACCCCTGACC-----	2290	
Qy	119	GlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValaAspProAsnAsnAsp	138	
Db	2289	GACGGCAAC-----GGCAACCGATTGGCCAGGTCAACGCC-----	2254	
Qy	139	AsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGly	158	
Db	2253	-----GACGGCAGCGGCAACTGG-----	2236	
Qy	159	AlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrlle-----AlaAsnPro	177	
Db	2235	---ACCTTACCCCGAGCAGCGCGTGTGCCAAC--GGCACGGTGGTCAACGGCCACCGCT	2182	
Qy	178	ThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLys	197	
Db	2181	ACCGACCGCTCCGGCAACGCGAGTTCGCGGCCAGCGTCAACGTGGACCCCGTGGCACCG	2122	
Qy	198	SerSerSerSerLeuGlyGlnAlaSerLysTrpIlePheValGlyValAlaAspPro---	216	
Db	2121	GCCACGCCAGTGTCAACCCGAGCAACGGCACCAAGCTCAGCGGCACCGCGCGCGCGC	2062	
Qy	217	-----AsnAsnProValPheTrp-----SerArgAspGly	226	
Db	2061	GCCACGTGACCCCTGGCCGATGGCAACGGCAATCCCATCGGGCAGGTCAACCGCCGATGGC	2002	
Qy	227	GlyAlaThrTrpGlnAlaValProGlyAlaPro-----	237	
Db	2001	AGCGGCACACTGGAGCTTTCATCCGACCAACGCCCTGTGGCCACCGCACCGTGTCAACGCC	1942	
Qy	238	-----ThrGlyPhe	240	
Db	1941	ACGGCCACCGACGCCTCGGCAACACCGATGCGGGCAGCAGTGTCAACCGTGGACTCGGTA	1882	
Qy	241	IleProHisLysGlyValPheAspProValAsnHisValLeuTyrlleAlaThrSerAsn	260	
Db	1881	GCCCGCGCCACGCCAGTGATCAACCCAGCAACCGCACCGCCACCGTCTCAGCGCACCGCGAG	1822	

Qy	261	ThrGlyGlyProTyr-----AspGlySerSerGlyAspValTrpIlysPheSerVal	277
Db	1821	CCGGCAGCAGCGTGCACCTGACCGATGGATGCAACAGCAACCCGATCGGCAGGTCAACCGCC	1762
Qy	278	Thr---SerGlyThrTrpThrArgIleSerProValPro-----	289
Db	1761	GACGGCAGCGGCAACTGGAGCTTCACCCCGTCCACGCCCTGGCGGATGGAACCGTGGTTC	1702
Qy	290	-----SerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAsp	306
Db	1701	AACGCCACGGCCACCGATCCGGCGGGCAACACGACGGCGCGGCGCAGCACCGTGCAT	1642
Qy	307	ArgGlnHisProAsnThrIleMetVal-----AlaThrGlnIleSerTrpTrp	322
Db	1641	GGCGTGGCGCGCACCACCGCGACCGTCAACCTGAGCAACGGCAGCAGCTCAGCGGCAT	1582
Qy	323	ProAsp-----ThrIleIlePheArg-----	329
Db	1581	GCGGAAACCGGGCAGCAGCGTGAATCTCACGACGGCAACGGCAATCGATCGCCGAGGTC	1522
Qy	330	SerThrAspGlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArg	349
Db	1521	ACCGCGACGGCAGCGCACTGGACC-----TACACCCCGT--CCACGCCGA	1476
Qy	350	SerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnPro	369
Db	1475	TCG-----CCAAACGGCACCGTGGTCAACGTGGTGGCCCGCAGGACG	1437
Qy	370	AsnProProValPro-----SerProLysLeuGlyTrpMetAspGluAlaMetAla---	386
Db	1436	CCGCGGCAATAGCAGCGCGGGCGCAGCGTCAACCGTGGAC--TCGAAAGCCCGCGCGGCT	1378
Qy	387	-----IleAspProPheAsnSerAspArgMetLeuTyrGlyThr-----GlyAla	401
Db	1377	CGGTGGTGTTCACCCGACCAACGGC--ACCACGCTCAGCGGCACCGCCGAGCGCGGCT	1321
Qy	402	ThrLeuTyrAlaThrAsn-----AspLeuThrIlysTrpAspSer	414
Db	1320	ACCGTAGCGCTGACCGACGGCAACGGCAACCCGATTGGCCAGGTCAACCGCC--GACGGC	1264
Qy	415	GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAsp	434
Db	1263	AGTGGCAACTGGAGCTTCACACCGGGCACCGCGTGGCCCAACGCAACCGTGGTCAACGCC	1204
Qy	435	LeuIleSerProProSerGly-----AlaProLeuIleSerAlaLeuGlyAspLeu	451
Db	1203	ACGGCCAGCGACCGGACCGGCAATACGAGCGCTCCGCGCAGCACCCAGCTGGAGTCGGT	1144
Qy	452	GlyGlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProVal	471
Db	1143	-----GCGCCGCGCGCGGTGGTCAATCCGACCAACGCAACGGAGTC	1105
Qy	472	PheThrThrGlyThrSerValAspTyrAlaGlnLeuAsnProSerIleIleValArgAla	491
Db	1104	GTCATCAGCGCACCC-----GCGCAACCGCGCGCCACCGTGCACCTGACCGCAT	1057
Qy	492	GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly	511
Db	1056	GGCAGCGGCAATCCGATCGGGCAG-----GTCACCGCGCGCAGCGC	1018
Qy	512	GlyLysAsnTrp-----Phe	516
Db	1017	ACGGGCACTGGAGCTTCACCCCGTCCACGCCGCTGGCGGATGGAAACGTTCAACGCC	958
Qy	517	GlnGlySerGluProGlyGlyValThrThrGlyGly-----ThrValAlaAla	532
Db	957	ACCGCTACCGACCCGGCGGC--AATACCGCGCGCGGCGCAGCACTACCGTGGACGCC	901
Qy	533	SerAla-----AspGlySerArgPheValTrpAlaPro	543
Db	900	ATCGCGCGCGCCACGCCGACCGTCAACTGAGCAATGGCAGCAGC---CTACGGCGCACT	844











```
QY 94 ProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrPro 113
Db 427012 AACAACTTGGC-----
QY 114 LeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaVal 133
Db 427000 -----ATGCCAACGGGGCAACTTCAACACCGGCGCATGGC 426965
QY 134 AspProAsnAsnAspAsnIleuTyrrPheGlyAlaProSerGlyLys--GlyLeuTrp 152
Db 426964 AACACCGGCAACAACATCGCTTGTTCACACCGGCAACAACATCGGCGCATGG 426905
QY 153 ArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThr 172
Db 426904 CTGACCGGCGAGCGGTCTCGGCTTACGCTCCCTGAACTCCGGCGCGGCAACACCGGT 426845
QY 173 TyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrp 192
Db 426844 TTCCTCAACTCGGCGACCGCAACACCGGC----- 426815
QY 193 ValAlaPheAspLysSerSerSerLeuGlyGln---AlaSerLysThrIlePheVal 211
Db 426814 ---TTGTTCAACTCGGCGACCGCAACACCGGCTTGTTCAACTCGGCGACCGCAACGTC 426758
QY 212 GlyValAlaAspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGln 231
Db 426757 GGCATCGGCAACATGGGCGACCGGCGCTTCGGCTCGGCGCTATCCGGCGACGACGAGTG 426698
QY 232 AlaValProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsn 251
Db 426697 GGCATCGGCGGCGCAACATCGGCGAGTTTC---AACATCGGCTTGTTCATCTCGGCGACC 426641
QY 252 HisValLeuTyrrIleAlaThrSerAsnThrGlyGlyProTyrrAspGlySerSerGly-As 271
Db 426640 GGCAATGTCGGCATCGCAACTCGGCGACCGGCAACGTCGGCATCGGCAACACCGGCGACC 426581
QY 271 pValTrpLysPheSerValThrSerGlyThr-----Tr 282
Db 426580 GGCAACACCGGCGATCGGAACACGCGGCAACTACAAACCGGCTTGCTCAACCGGCGCGCTG 426521
QY 282 p-----ThrArgIleSerProValProSerThrAspThrAlaAsnAs 296
Db 426520 GTCAACACCGGCGATCGCAACCGGCGCAACACACACCGGCGCTTCAACATCGGCGACC 426461
QY 296 pTyrrPheGlyTyrrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAl 316
Db 426460 TTCAACACCGGCGATCGCAACCGGCGCACTACAAACCGGCTCTCAACACACCGGTAGC 426401
QY 316 aThrGlnIleSerTrp---TrpProAspThrIleIle---PheArgSerThr----- 331
Db 426400 TACAACACCGGCGATCGGAACACCGGCGAGACTACGCGCGCGCTTATCAACCGGCGAGC 426341
QY 332 -----AspGlyGlyAlaThrTrpThrArgIleTrpAspTrpTrpSer-- 345
Db 426340 ATGAACAACGGGTGCTCTCGGCGCGCGACCGGCGAGGCGCTGCTCGGCGCGCACTACACC 426281
QY 346 -----TyrProAsnArgSerLeuArgTyrrValLeuAspIleSerAlaGl 360
Db 426280 ATCAACATCGAGCGACCTCGCGGTCTCTCAATGTCGACATCCCGGTCAACATCCCCATC 426221
QY 360 uProTrpLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTr 380
Db 426220 ACCG---GCGACATCAACATGTCTCCATCCCGCGCATTTCTTCCCGAGA-ATC----- 426170
QY 380 pMetAspGluAlaMetAlaIleAsp---ProPheAsnSerAspArgMetLeuTyrrGlyTh 399
Db 426169 ---GACCGCGGAGCGGACGTCGATAGGCATTCCTGAGCGCACCGCTCTTGCGCCCGGT 426114
QY 399 rGly---AlaThrLeuTyrrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnI 418
Db 426113 CGGTCCGATCACCTCTGCATGGCGGCGACGCGTCCGCGCCCGCTTGGCGCCCGGT 426054
```



```
US-09-902-540-1027/c
; Sequence 1027, Application US/0902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1027
; LENGTH: 10317
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-1027

Alignment Scores:
Pred. No.: 5,05e-06 Length: 10317
Score: 184.00 Matches: 155
Percent Similarity: 29.41% Conservatives: 70
Best Local Similarity: 20.28% Mismatches: 228
Query Match: 4.56% Indels: 312
DB: 4 Gaps: 40

US-09-917-376-3 (1-740) x US-09-902-540-1027 (1-10317)
QY 44 TrpAspAlaAlaAsnGlyArgTrpIlePro----- 53
Db 4722 TGGGATGCGGAGCGGAGGCTTCATCATGACAGCGGACGGGCTCGGTGCAAAACCG 4663
QY 54 -----LeuLeuAspTrpValGly----- 61
Db 4662 CCCACGGGGTGTCCTCAATGACACGGTGGGCTCACCTCGTGGCAGCGGACCTGGGAG 4603
QY 62 AsnTrpGlyTyraAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysVal 81
Db 4602 GAGTGGGCGGTCGGGAGACTTCGGGCGCGGCGGCTCCGGCGGGTTCGGTGGCGCC 4543
QY 82 TrpAlaAlaValGly----- 86
Db 4542 TTCAGATATCGGGAGCGGCATCAGCGTGCTCTTCGGTGGAGAGAGCTGCACGGGGTGC 4483
QY 87 MetTyraAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGlnGly 106
Db 4482 AAGCTGTCGGACACGTGGCAGTTCAACTCAGCCACT----- 4447
QY 107 AlaThrTrpGlnIleThrProLeuProPheLysLeuGlyAsnMetProGlyArgGly 126
Db 4446 CGGACTGTCAG-----CAGTGGGGGCGCGGGGCGCTCCGGCGCT 4405
QY 127 MetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyraPheGlyAlaPro 146
Db 4404 TCGGGGCGGGGTGACGCGCTCCCGGTCTCGGCGCATCTTGCTGTTCGAGGTACC 4345
QY 147 SerGlyLysGly-----LeuTrpArgSerThrAsp 156
Db 4344 TCGGGAGCGGGCTCGCAGGAGGCGCTTTCGTCAACACGAGCTTTGGAGCCATGCCCCAG 4285
QY 157 SerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyraIleAlaAsn 176
Db 4284 GGGTGGGCCAAGTTGCTCCACCGGGAGCTTCCCGCGCGCGCGCTTCGCGCGCGCC 4225
QY 177 ProThrAspThrThrGlyTyraGlnSerAspIleGlnGlyValValTrpValAlaPheAsp 196
Db 4224 CGGCTGACACCGAGC-----TCCGGGCTCATGTGGATCTCGCGCGGT 4183
QY 197 LysSerSerSerSerLeuGlnAlaSerLysThrIlePheValGlyValAlaAspPro 216
Db 197 LysSerSerSerSerLeuGlnAlaSerLysThrIlePheValGlyValAlaAspPro 216
```

```
Db 4182 GCGACGCGCGCGCCCATCGGC-----GCCAGCCTG 4153
QY 217 AsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAla 236
Db 4152 GCGACGTGCTCCACGTACACCGCTCGACCAATGCC---TGGCGAGCGCGCCCTCGCTG 4096
QY 237 ProThr-----GlyPheIleProHisLysGlyValPheAspPro 249
Db 4095 CCGCGGCTCGGTGGTGGATTGATGCTTATCGCGCGCGCGGATGTAT----- 4042
QY 250 ValAsnHisValLeuTyraIleAlaThrSerAsnThrGlyGlyProTyraAspGlySerSer 269
Db 4041 -----CTCTTCGGAGAACGTCAC-----GGTGGCGCGCGC 4009
QY 270 GlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArgIle----- 285
Db 4008 AACGATCTGCTCCGGTTCAGC-----GCGGGGCGTGGACACGCTCACCAGCGCGCGC 3955
QY 286 -----SerProValProSerThrAspThrAlaAsnAspTyraPheGlyTyraSerGly 302
Db 3954 GCACGCGGGGAGTCTCT---CCGAGCACT-----GGTGGG 3925
QY 303 LeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrp 322
Db 3924 GTGAAGCTCGAGACGGAC---CCACAGACA----- 3898
QY 323 ProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrpAsp 342
Db 3897 GCGACCTCTCTCTCTCACGGGAC-----GGGCGGTGTGGAC-----TTGCAC 3850
QY 343 TrpThrSerTyraProAsnArgSerLeuArgTyraValLeuAspIleSerAlaGluProTrp 362
Db 3849 GGAACGCGGTGGCAGCGGGCCACT-----TCGTCGTCTCGCGCGGAGACC 3805
QY 363 LeuThrPheGlyValGlnProAsnProValProSerProLysLeuGly----- 379
Db 3804 CTCACGCTCGGGGACGATTTCGGCGGAGTGACGAGCGGCTTCAGTGGCGCCACCATC 3745
QY 380 TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyraGlyThr 399
Db 3744 TGGGTG-----GTCGCGCCACCGCGCTTCAGCGCGACCATCGTTCTACGTC 3700
QY 400 GlyAlaThrLeuTyraAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHis 419
Db 3699 GGTCTGACG-----GGTGG----- 3685
QY 420 IleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAspLeuIleSerProPro 439
Db 3684 -----GTCCGACGCTACAGCTCAGCGGGGTGCGC 3655
QY 440 SerGlyAlaProLeu-----IleSerAlaLeuGlyAspLeuGlyGly-----PheThrHisAla 457
Db 3654 GCGGGGTTCGCTCTCCGTCTATGCTACTACGAGTCGGGAGGCGCTTATTCACAAAG 3595
QY 458 AspValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSer 477
Db 3594 GAGCTGGGGAGTGGGCGCGCTGGTC-----GCGACACGACG 3556
QY 478 ValAspTyraAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSer 497
Db 3555 TTGGACATCGC----- 3544
QY 498 SerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGln 517
Db 3544 ----- 3544
QY 518 GlySerGluProGlyGlyVal-----ThrThrGlyGlyThrValAlaAlaSerAla 534
Db 3543 ---CTGGAAACCGAGCCCTCGGCGACACCGACGACGACCCCTCGTCCCTCCCGCG 3487
QY 535 AspGlySerArgPheValTrpAla----- 542
Db 3486 GATTGGCCCGCTTCATCAATGCTCGCGGTATGCGCGGGGTTCGCTCCAGTCAGGACTT 3427
```



```
QY 543 ProGlyAspProGlyGlnProValTyrAlaValGlyPheGlyAsnSerTrpAlaAla 562
DB 3426 CCGCGCAGCCCAATGGCCCCCAGAGAGCAGCGCGCGCATCGTGAGGAGGTCTCCGCG 3367
QY 563 SerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPhe 582
DB 3366 ACCTAC--ATCCCTGCGTCCGCA-----CCGCCACTCAG 3334
QY 583 TyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnPro 602
DB 3333 GCCGTCGCTCGAGCGCAGGTC----- 3310
QY 603 ValAlaAlaGlyLeu-----ProSerSerGlyAla 612
DB 3309 GTGGGGGCATCGCTCGCAGAGCGTTGACGACCTGGGTCTACAATCCTCCCGCGGGTGT 3250
QY 613 ValGlyValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaSer 632
DB 3249 GTCCGCGCGGTG-----GCGCTACG----- 3229
QY 633 SerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSer 652
DB 3228 -----GATGGCATCCGGGCGGTG--- 3211
QY 653 SerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPhe 672
DB 3210 -----GCGCGGGGCGGTATCAGCTCACC CGCGCGCCGCCACCGG 3181
QY 673 ValValGlyThrIleGlyGlyValThrGlyAlaTyr----- 684
DB 3180 ACGGTGCCGACGGTC-----GCGGAGGGGCGGTATCAGCTCACC CGCGCGCCGCCACG 3127
QY 685 -----ArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAspAsp 699
DB 3126 GAGTTGATCTCGTCGTCGAGCGCGGAGGATGTCTGGGACTGGGTGCATCTCTCG 3067
QY 700 GlnHisGlnTyrGly 704
DB 3066 CAGGCACAGCGGGC 3052
```

## RESULT 7

```
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: PRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2
```

## Alignment Scores:

Pred. No.:	0.067	Length:	4403765
Score:	184.00	Matches:	204
Percent Similarity:	32.65%	Conservative:	116
Best Local Similarity:	20.82%	Mismatches:	306
Query Match:	4.56%	Indels:	361

```
DB: 3 Gaps: 50
US-09-917-376-3 (1-740) x US-09-103-840A-2 (1-4403765)
QY 14 GlyGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeu--- 32
DB 3745039 GCGCGTGGCGGTCAACCGGATGGC-----GATGCGGGCGGTGGCGGTGTCGAATCT 3744989
QY 33 -----TyrValArgThrAspIleGly 39
DB 3744988 GTTGGGCGAGAACCGCGCAGCATCGCGGCGCTCGAGGCCGAGTACGAGTTGATGTGGGC 3744929
QY 39 ----- 39
DB 3744928 CGCGCATGTGGCGCGATGGCGCGGTACCATTTCCGCGCGCTCGGTGTCGCCGCGCGT 3744869
QY 40 -----GlyMetTyrArgTrpAspAlaAsnGlyArg----- 50
DB 3744868 GCCGCGTTCAGCCACCGCGCAGG-CGTTGGGGGTGTGTCTGCGCGGTTCCTCAATG 3744810
QY 51 -----Trp---IleProLeu 54
DB 3744809 CTCATTTCGCGACCGCGAAGATGTTGAGGCTTAACCGCGGCTTGGCAATGTCGTA 3744750
QY 55 LeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAsp 74
DB 3744749 ATTACACGTCGGTGGGCAATGTCGGATAT-----TCAACCTGGCGCGAG 3744702
QY 75 ProIleAsnThrAsnLysValTrp-----AlaAlaValGlyMetTyr 88
DB 3744701 CCAATGTCGGTGGCGAATTTGGGTGTCGCCAACCGCGGTAGCGGAATTCGGTTTCG 3744642
QY 89 ThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArg----- 101
DB 3744641 GCATATCGGCAACCGCAACTTCGGGTTCGGCAACTCGGGTCTTGGGTGCGCCCGGCA 3744582
QY 102 -----SerSerAspGlnGlyAlaThr 108
DB 3744581 TGGCAATATTGGGTGGGCAATGGCGCAGCAGCAACTACGCGCTCGCAACCTGGGTG 3744522
QY 109 TrpGlnIleThrProLeuProPheLysLeu-----Gly 119
DB 3744521 TGGCAACATCGGTTTTGCACACCGGTAGCAACACATCGGATCGGTTGACCGGG 3744462
QY 120 GlyAsnMetProGly-----ArgGlyMetGlyGluArgLeuAlaValAsp 134
DB 3744461 AC-AACTGACTGGCATTTGGGGCGCTGAATTCAGGAACCGGTAACTGGGGTTGTTCAAC 3744403
QY 135 ProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLys---GlyLeuTrpArg 153
DB 3744402 TCCGCGACCGGCAACATCGGTCTTCAATTCGGGACCGGCAACTTCGGGGTTATCAAC 3744343
QY 154 Ser-----ThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAsp 169
DB 3744342 TCGGGCAGCTACACACCGGTGTGTAATGCGGGAGCGGCAGTACCGGGTTGTTCAAC 3744283
QY 170 Val-----GlyThrTyrIle 174
DB 3744282 GTTGGTGGTTCACACCGGGTGTGCCAACGTTGGGTAGTATAACACGGGAGCTTCAAC 3744223
QY 175 AlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTyrValAla 194
DB 3744222 GCGGCAACACCAATACGGGTGGCTTCAACCGGGGCAACGTCACACCGGCTGGCTGAAC 3744163
QY 195 PheAspLysSerSerSerSerLeu-----GlyGlnAlaSerLysThrIlePheVal 211
DB 3744162 ACCGCAACACCAACCGGCATCGCAACTCGGCAATGTCAACACCGCGCGGTTCATC 3744103
QY 212 GlyValAlaAspProAsnAsnProValPheTrpSerArgAsp----- 225
DB 3744102 TCG---GGCAACTTCAGCAACCGGTGTGCTGTGGCGGGGTGACTACGAGGCGCTGTGGGG 3744046
QY 226 -----GlyGlyAlaThrTrpGlnAlaValPro-----GlyAlaProThrGlyPhe 240
```







Pred. No.: 1.05e-05 Length: 11679  
 Score: 181.50 Matches: 204  
 Percent Similarity: 31.05% Conservative: 90  
 Best Local Similarity: 21.54% Mismatches: 344  
 Query Match: 4.50% Indels: 309  
 DB: 4 Gaps: 54

US-09-917-376-3 (1-740) x US-09-328-352-1377 (1-11679)

QY 7 ThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGly----- 21  
 DB 1627 ACGTGGACA--GTAAGTGTGCGGGTAGTGGTTGGTGTGATGCGAGATAAACAGATT 1683  
 QY 22 -----IleValPheAsnGluGlyAlaProGlyIleLeuTyValArg---ThrAsp 37  
 DB 1684 GATGCTAAAGTAACGTTTACAGATCGACGAGGTAAATACAGCACTGTTAAAGTACGCAA 1743  
 QY 38 IleGlyGlyMetTyArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrp 57  
 DB 1744 ATT-----TATACATTAGACACAGCT--GCTCTGCGAGCCAGTAATCGACCCA 1791  
 QY 58 ValGlyTrpAsnAsnTrpGlyTyAsnGlyVal-----ValSerIleAlaAspPro 75  
 DB 1792 GTT-----AACGGGACAGACCCCAATTACAGGTACAGCAGAACCT 1830  
 QY 76 IleAsnThrAsnLysValTrpAlaAlaValGlyMetTyThrAsn----- 90  
 DB 1831 GGTTCAACAGTAACAGTA-----ACCTATCTTAATGGTGACACAGCAACA 1875  
 QY 91 -----SerTrpAsp---ProAsnAspGlyAlaIleLeuArgSer 102  
 DB 1876 GTTGTAGCAGACCGGCGAGTTGGTCAGTACCACCAACCCCTGGC-----CTTAATGAT 1929  
 QY 103 SerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsn--- 121  
 DB 1930 GGTGACGAAGTTGAGGCAATTGCTACAGATCCA-----GCAGGCAACCCA 1974  
 QY 122 ---MetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsn----- 137  
 DB 1975 TCTTTGCGAGGTACAGCTACTGTTGAT-----GCAGTTGGTCCAAATACCGATGGTGT 2028  
 QY 138 -----AspAsnIleLeuTyPheGlyAlaProSerGly 148  
 DB 2029 AACTTTACGGTTGATTACGTACACGCTGCAATGTGATTAAATGCATCAGAACGCTGACGC 2088  
 QY 149 Lys-----GlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMet 164  
 DB 2089 AACGTTACTGTTACTGTTGATTGAAACCGTTCCGCGAGATGCAGCAAAATACAGTGGTC 2148  
 QY 165 ThrAsnPheProAspValGlyThrTyIleAlaAsnProThrAspThrGlyTyGln 184  
 DB 2149 ACTGTTGTAATCAATGGCGCAGACGTATCTGCACTGATAGTAGACACAGCAGGC----- 2202  
 QY 185 SerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGln 204  
 DB 2203 -----ACATGGACAGTAGCCGTACCAGGTAGTACTTAACCTGCCGAT 2244  
 QY 205 AlaSerLysThrIle----- 209  
 DB 2245 GCAGATAAGACGATTGCTAAAGTAAAGTTTACAGATGCGCGAGGTAATAGCAGCAGT 2304  
 QY 209 ----- 209  
 DB 2305 GTTAAACGATACACAAACATATACAAATCGATACCACTGCACCTGTATGCGACAGTAATTAAC 2364  
 QY 210 -----PheValGlyValAlaAspProAsnAsnProValPheTrp 222  
 DB 2365 CCGGTTAAACGGGACAGACCGGATTACAGGTACGGCAGACGCTGGTTCAACAGTAACTGTG 2424  
 QY 223 SerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaPro---ThrGlyPheIle 241  
 DB 2425 ACTTATCCAGATGGCAGTACACACACAGTGTGTTGCGAGGACCGGATGGCACTTGGACAGTA 2484

QY 242 ProHisLysGlyValPheAspProValAsnHisValLeuTyIleAlaThrSerAsnThr 261  
 DB 2485 CCAAAACCCAGGTTTAAATGATGGC---GATAAAGTTTACAGCAATGCTACAGATCCAGCA 2541  
 QY 262 GlyGlyProTyArgTrpAspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThr 281  
 DB 2542 GGCACACCA-----TCATTACCAGGTACA 2565  
 QY 282 TrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyPheGlyTySer 301  
 DB 2566 GCTACTGTTGATGCGAGTGGTCCAAATACCAATGCTGTTAAT-----TTCACGGTTGAT 2619  
 QY 302 GlyLeuThrIleAspArg----- 307  
 DB 2620 TCAGTAAACAGCTGACAATGTGATTAAATGCATCAGAAGCATCAGGCAACGTTACTGTTACA 2679  
 QY 308 -----GlnHisProAsnThrIleMetValAlaThrGlnIle 319  
 DB 2680 GGTGTATTGAAAAACGTTCCGCGCAGATGCAGCAATATACAGTG---GTCACTGTTGTGATC 2736  
 QY 320 SerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpTrpArg 339  
 DB 2737 AATGCCAGACGTATACGTGCAACTGTAGTAGACACAGCAGGC-----ACATGGACA--- 2787  
 QY 340 IleTrpAspTrpThrSerTyProAsnArgSerLeuArgTyValLeuAspIleSerAla 359  
 DB 2788 -----GTAAGCGTACCAGGTAGTAGTCTGACTGCGGATGCAGATAAGACGATT 2835  
 QY 360 GluProTrpLeuThrPhe----- 365  
 DB 2836 GATGCCAAAGTAACGTTTACAGATGCAGCGGTAAACAGCAGCAGTGTAAACGATACACAC 2895  
 QY 366 -----GlyValGlnProAsnProProValProSerProLysLeuGlyTrp 380  
 DB 2896 ACATATACAGTTGATACGTTGCAACCAATGCACCCGGTG----- 2934  
 QY 381 MetAspGluAlaMetAlaIleAspProPheAsnSerAsp----- 393  
 DB 2935 CTTGATCCGATCAATGCAACAGACCCAGTGAGCGGTGAGCAGAGCCTGGTTCAACAGTG 2994  
 QY 394 ArgMetLeuTy-----GlyThrGlyAlaThrLeuTyAlaThrAsnAsp----- 408  
 DB 2995 ACTGTGACTTATCTGTATGGCACCACCTGCAACAGTGGTAGCAGCAGCCGATGTTAGTGG 3054  
 QY 409 -----LeuThrLysTrpAsp 413  
 DB 3055 TCAGTACCAACCCAGGTAACCTGGTGGATGGGATACAGTACAGTCAACAGCAACTGAC 3114  
 QY 414 SerGlyGlyGlnIleHisIle-----AlaPro 422  
 DB 3115 CCTGACGCAACACTTCATTGCGAGGTACAGGCACAGCTTTCAGCAGACATCAGCACCT 3174  
 QY 423 MetValLysGlyLeuGluGluThrAlaValAsnAsp-----Leu 435  
 DB 3175 GTGGTT---GCGCTGGATGACGTTGTGACGAATGACAGCACACAGCACTTACGGGTACA 3231  
 QY 436 IleSerProProSerGlyAlaProLeuIleSerAla----- 447  
 DB 3232 GTGAACGATCCGACACCCACTGTGTTGTCATGTTGGATGGCGTTGACTATCCGCGAGTG 3291  
 QY 448 ---LeuGlyAspLeuGlyPheThrHisAlaAsp-----ValThrAla 461  
 DB 3292 AACAAATGGTAC---GGCACCTGGACGCTTGACAGACAAATACACTCTCTCGGTTAACTGAT 3348  
 QY 462 ValProSerThrIlePheThrSerProValPheThrThrGly----- 475  
 DB 3349 GGTCCACACACCACTTACCGTGACTGCAACAGATGCAGCAGGCAATCAGGTACAGATACA 3408  
 QY 476 -----ThrSerValAspTyAlaGluLeuAsnProSerIleIleValArgAlaGlySer 493  
 DB 3409 GCAGTGGTGCAGCATTTGATACACAGCAACCAATGCACCGGTACTTGTATCCGATCAATGCG 3468  
 QY 494 PheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyLys 513











OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-17474

Alignment Scores:  
Pred. No.: 0.000121 Length: 29927  
Score: 177.00 Matches: 189  
Percent Similarity: 28.82% Conservative: 56  
Best Local Similarity: 22.24% Mismatches: 292  
Query Match: 4.39% Indels: 314  
DB: 4 Gaps: 48

US-09-917-376-3 (1-740) x US-09-949-016-17474 (1-29927)

QY 14 GlyGlyGlyPheValAspGly-----IleValPheAsnGluGly 27  
DB 19367 GGGGGCTCTGGGAGCGGAGCGCTGGCTCCCTCTTCCACCGTTTTATTCACAGGGGAC 19426  
QY 28 AlaProGlyIleLeuTyrValArgThr-----AspIleGlyGlyMet 41  
DB 19427 AGGCTGGGGATTTGTAATTTGGGCGCGTGTGGCTGAGGCTGAGGGACTT-GGGGGG--- 19482  
QY 42 TyrArgTTPAspAlaAlaAsn----- 48  
DB 19483 TGGCGGTGGGAGCGCGGAGGTATAAAGTATATAATCATAGTAAACAACTCAGAAATG 19542  
QY 49 -----GlyArgTTPileProLeu-LeuAspTTPValGlyTTPAsnAsnTTPGlyTyrAs 66  
DB 19543 GACCCCGAGCGGTGGTGGCGGTAGCTCTCCAGCTCTCCCTGGCCCGAGGCGGAGGAGA 19602  
QY 66 nglyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTTPAlaAlaValGI 86  
DB 19603 GGGGTCGCGATCCCTCCGCGGTCTCTCTC----- 19633  
QY 86 yMetTyrThrAsnSerTTPAspProAsnAspGlyAla-----IleLeuArgSerSe 103  
DB 19634 -----CTGGGTACTGCGCTTGGGTGGGAGACGAGCCTACTTCTTGATCCGTCCTT 19686  
QY 103 rAspGlnGlyAlaThr-----TTPGlnI 111  
DB 19687 TGCCGAGCGGGGACCCAGTAAATTAGCCGTTGGAGCCCGCAGCGCTGCTGCTGTTG 19746  
QY 111 eThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGly----- 128  
DB 19747 CGCACCGAGTCTTGG-----GGACCTGGTGTCTCCCGGAGAACTTGGGGACCTGGTA 19800  
QY 129 -----GluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyr----- 142  
DB 19801 TCCCCGGGAGAGCTTGGGACCTGGTGTCCTCCGGGAGAGCTTGGGTACTGTTTCTCT 19860  
QY 143 -----PheGlyAlaProSerGlyLysGlyLeu----- 151  
DB 19861 GGAAGAGCTTGGACACCTGGTCTCTGGGAGGCGCTTGGGACCTGGTGTCTCTGGGAGA 19920  
QY 152 -----TTPArgSerThrAspSerGlyAlaThrTTPSerGlnMetThrAsnPheProaspVa 170  
DB 19921 GGCTTGGAGATCTGTGTCTCTGGGAGAGGCTTGGGGA----- 19957  
QY 170 lGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyVa 190  
DB 19958 -----CTGGGTCTCTCTGGAGA 19974  
QY 190 lValTTPValAlaPheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIlePh 210  
DB 19975 GGCTTGG-----GGACCTGGTGACCTTGGAGAGCTTGGAGACCTGGTGT 20019  
QY 210 eValGlyValAlaAspProAsnAsnPro-ValPheTTPSerArg----- 224  
DB 20020 TCTGGGAGAGGCT-----TGGGAGCCTGGTGTCTGGGAGAGGCTTGGGGACCTGGTGT 20073  
QY 225 -----AspGlyGlyAlaThrTTP-----GlnAlaValProG 235  
DB 20074 CTCTGGAGAGGCTTGGACACCTGGTGACCCGGAGGGGCTTGGGGATCTGGTGTCTCCGG 20133

QY 235 lyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeuT 255  
DB 20134 GAGAGCCTTGGGGA-----CTGGTGTCTCTG----- 20159  
QY 255 yrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTTPlysp 275  
DB 20160 -----GGAGAGGCTTGGGACCT-----GGTGACCTTGGAGAGCTTGGGAC 20202  
QY 275 heSerValThrSerGlyThrTTPThrArgIleSerProValProSerThrAspThrAlaA 295  
DB 20203 CTGGTGTCTCTGAGAGAGCCTTGGGGA-----TCTGGTGTCTCCAGGAGAGCTTGGGAC 20256  
QY 295 snAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetV 315  
DB 20257 CT-----GGTGTCTCTGGAAGAGCTTGGAC----- 20282  
QY 315 alaAlaThrGlnIleSerTTPTrpPro-----AspThrIleIlePheArgS 330  
DB 20283 -----ACCTGGTGTCTCTGGGAGAGGCTTGGGACCTGGTG----- 20318  
QY 330 erThrAspGlyGlyAlaThrTTPThrArgIleTTPAsp-TTPThrSerTyrProAsnArg 349  
DB 20319 -----TCTGGGAGAGGCTTGGGACCTGGTGT-----CTTGGGAGA 20355  
QY 350 SerLeu-ArgTyrValLeuAspIleSerAlaGluProTTPLeuThrPheGlyValGlnPr 369  
DB 20356 GCGTTGGAGA-----TCTGGTGGAGCCGGGAGAGGCTTGGGGA----- 20392  
QY 369 oAsnProProValProSerProLysLeuGlyTTPMetAspGluAlaMetAlaIleAspPr 389  
DB 20393 -----CTGGTGTCTCCGGAGAGGCTTGGGACCTGGTGTCTCCGGGAGAGGCTTGGACACC 20448  
QY 389 o-PheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspL 409  
DB 20449 TGGTGTCTCCAGGAGAGGCTT-----GGGACCTGGTGACCTTGGAGAGGCTGGGAC 20502  
QY 409 euThrLysTTPAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluG 429  
DB 20503 TGTGACCCGGGAGAGCCTTGGG----- 20525  
QY 429 luThrAlaValAsnAspLeuIleSer-----ProProSerGlyAlaProLeuIleSerA 447  
DB 20526 -----GACCTGGTGTCTCTGGGAGAGCCTTGGGAGCCT----- 20558  
QY 447 laLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSerThrIleP 467  
DB 20559 -----GGTGACCTTGGAGAGGCTTGGGACCTGGTGTCTCGGGAGTGCCTTGGGACCT- 20612  
QY 467 heThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerI 487  
DB 20613 --AGTGACCCGGGAGAGGCTTGGGACCTGGTGTCTCCGGGAGAGGCTTGGGACCTGGTG 20670  
QY 487 leIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHisValAlaP 507  
DB 20671 TCCTG-----GGAGAGCCTTGGGATCTGCTGTCTCTG- 20702  
QY 507 heSerThrAspGlyGlyLysAsnTTPPheGlnGlySerGluProGlyGlyValThrThrG 527  
DB 20703 -----GGGAGAGGCTTGGGAGCCTGGTGTCTCTCGGAGAGAGCCTTGGGAGCCTGGTGACC 20757  
QY 527 lyGlyThrVal---AlaAlaSerAlaAspGlySerArgPheValTTPAlaProGlyAspP 546  
DB 20758 CGGAGAGGCTTGGACACCTGGTGTCTCCGGGAGAG-----GCTTGGGACCTTGGTGACC 20811  
QY 546 roGlyGln-----ProValValTyrAlaValGlyPheGlyAsnSerTTPAlaAlaAs 563  
DB 20812 CGGAGAGCCTTGGGAGCCTGGTGTCTCTGGGAGAGGCTGGGGG---ACCTGGTGTCTCG 20868  
QY 563 erGln-----GlyValProAlaAlaGlnIleArgSerAspArgValA 578  
DB 20869 GGAGAGAGCCTTGGGAGCCTGGTGACCCGGGAGAGGCTTGG----- 20909  
QY 578 snProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGly---- 596



		Query Match:	4.39%	Indels:	314
		DB:	4	Gaps:	48
		US-09-917-376-3 (1-740) x US-09-949-016-17475 (1-29927)			
QY	14	GlyGlyGlyPheValAspGly-----lleValPheAasnGluGly	27		
Dd	19367	GGGGCTTCGCAGCGCCGACGCGTCCCTTTCAACCCTTTATTTCCAAAGGGAC	19426		
QY	28	AlaProGlyIleLeuTyrrValArgThr-----AspileGlyGlyMet	41		
Dd	19427	AGGTGGGGATTGATTTGGCGCGTGTTGGCTGAGGGTGACGGACTT-GGGGG-	19482		
QY	42	TyrArgTrpAspAlaAlaAasn-----	48		
Dd	19483	TGGCGTGGGAGCGCGAAGGTATAAACCGTATAAATCATAACTATAACAATCAGA	19542		
QY	49	-----GlyArgTrpIleProLeu-LeuAspTrpValGlyTrpAasnAsnTrpGly	66		
Dd	19543	GACCCCAGCGCTGGTCGCCGTAGCTCTCCAGCTCTCCCTGGCCCCAGGCCGAG	19602		
QY	66	nGlyValValSerIleAlaAlaAspProfileAasnThrAsnLysValTrpAla	86		
Dd	19603	GGGTTCGCATCCCTCCGCGGTCTCTCTC-----	19633		
QY	86	yMetTyrrThrAsnSerTrpAspProAasnAspGlyAla-----lleLeuArgSerSe	103		
Dd	19634	-----CTGGGTACCTGGCTTGAGGTGGGGGAACGAGCTACTTCTGTACCGTCTTT	19686		
QY	103	rAspGlnGlyAlaThr-----TrpGlnIl	111		
Dd	19687	TGCCGAGCGCGGACCCAGTGAATTAGCGCTTGGAGCCCGCACGCTCGCTGGCTTG	19746		
QY	111	eThrProLeuProPhelysLeuGlyGlyAsnMetProGlyArgGlyMetGly-----	128		
Dd	19747	CGACCGGAGCTTTGG-----GGACCTGGTGTCCTCCGGGAAAACCTTGGGACCTGTA	19800		
QY	129	-----GluArgLeuAlaValAspProAasnAsnAspAasnIleLeuTyr-----	142		
Dd	19801	TCCCCGGGAGAGGCTTGGGACCTGGTGTCCTCGGAGAGGCTTGGGTACCTGGTTCTCT	19860		
QY	143	-----PheGlyAlaProSerGlySylsGlyLeu-----	151		
Dd	19861	GGAAAGAGCTTTGGACACTGGTGTCCTGGAGAGGCTTGGGACCTGGTGTCCTGGGAGA	19920		
QY	152	---TipArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspVa	170		
Dd	19921	GGCTTGAGATCTGTTCTCTGGGAGAGGCTTGGGA-----	19957		
QY	170	lGlyThrTyrrileAlaAsnProthrAspThrThrGlyTyrrGlnSerAspiledinglyva	190		
Dd	19958	-----CCTGGTGTCCTGGAGA-----	19974		
QY	190	lValTrpValAlaPheAspysserSerSerSerLeuGlyGlnAlaserlysThrilleph	210		
Dd	19975	GSCTTG-----GGACCTGGTAGCTTTGGAGAGGCTTTGGAGACCTGGTG	20019		
QY	210	eValGlyValAlaAspProAasnPro--ValPheTrpSerArg-----	224		
Dd	20020	TCTGGGAGAGCT-----TGGGAGCTTGTTGTTCTGGAGAGGCTTTGGGAGCTGGTG	20073		
QY	225	-----AspGlyGlyAlaThrTrp-----GlnAlaValProG	235		
Dd	20074	CTCTGGAAGAGCTTTGGACACCTGGTGACCCGCGGAGGCTTTGGGATCTGGTGCTCCCG	20133		
QY	235	lyAlaProThrGlyPheIleProHisisLygilyValPheAspProValasnhisVallaut	255		
Dd	20134	GAGAGCTTTGGGA-----CCTGGTGTCCTG-----	20159		
QY	255	yrIleAlaThrSerAsnThrclglyProtyrAspGlySerSerGlyAspValTrplysp	275		
Dd	20160	-----GGAGAGCTTTGGGAGCT-----GGTGACCTTTGGAGAGGCTTTGGGAC	20202		

Db	20910	 -----ACACTGGTGTC	20922		
QY	597	--GlyValThrPheGlnProValAlaAlaglyLeupro----SerSerGlyAlaValGlyv	615		
Dd	20923	CGGGAGAGGCTTTGGAGCTTGGTCCCGGAGAGCTTTGGGACCAAGGTGACCTTGGAG	20982		
QY	615	aIMetPheHisAlaValProGlyLyseGlyAspleuTrpLeuAlalaSerSerGlyL	635		
Dd	20983	AGGCTTTGGGA-----CCTGGTGATCTTTGGAGAGGCTTTGGGAGCTTGCTCGGA-	21035		
QY	635	euTyrrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlav	655		
Dd	21036	-----GAGGTTACGGGGGCTGGTTGGGGGAGAACGTTGTGAGCCAAAAAGTCCC	21084		
QY	655	alAasn-----	656		
Dd	21085	TGAATCCCTCGNAGAGAGCGCATCGGAGCTCCCCCTGAGGGCGTTCATTTTGGRACC	21144		
QY	657	-----ValGlyPheGlyLysSerAlaProGlySerSerTyrrProAlav	671		
Dd	21145	CCCCTCCATGCGCTTTGAGGAGCTGTTTCGATTCCCTGGCCGGCT---CCCGCGG	21201		
QY	671	alphelvalValGlyThrIleGlyValThrGlyAlatyrrArgserAspaspCySGlyt	691		
Dd	21202	ATGCATCCAGTGCGACGCCAATTTCTGGCCAGGGGGAAGGAGGAAGCGCGGTGCGGG	21261		
QY	691	hrThrTrpValleulleAasnAspGlnHisGlnTyrrGlyAsnTrpGlyGlnAlailert	711		
Dd	21262	TGTC-----	711		
QY	711	hrGlyAspHisAlaAsnLeuArgArgValtyrileGlyThrAsnglyArgGlyIlevalt	731		
Dd	21274	CTGAGAGAGGGGCGCGCTCCCTAGG-----GGAGAGAGGCGACGTTGGGGGTTTC	21324		
QY	731	yrGlyAspileGlyGlyAla	737		
Dd	21325	CGGGGGCGCGGGCGGAGCA	21344		
RESULT 11					
US-09-949-016-17475 ; Sequence 17475, Application US/09949016					
; Patent No. 6812339 ; GENERAL INFORMATION:					
; APPLICANT: VENTER, J. Craig et al.					
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF					
; FILE REFERENCE: CL001307					
; CURRENT APPLICATION NUMBER: US/09/949,016					
; PRIOR FILING DATE: 2000-04-14					
; PRIOR APPLICATION NUMBER: 60/241,755					
; PRIOR FILING DATE: 2000-10-20					
; PRIOR APPLICATION NUMBER: 60/237,768					
; PRIOR FILING DATE: 2000-10-03					
; PRIOR APPLICATION NUMBER: 60/231,498					
; PRIOR FILING DATE: 2000-09-08					
; NUMBER OF SEQ ID NOS: 207012					
; SOFTWARE: FastSeq for Windows Version 4.0					







```
QY 28 AlaProGlyIleLeuTyrValArgThr-----AspIleGlyGlyMet 41
Db 19427 AGCTGGGGATTGATTGGCGCGGTGGTGGCTGAGGGTGCAGGACTT-GGGGG- 19482
QY 42 TyrArgTyrAspAlaAlaAsn----- 48
Db 19483 TGGCGGTGGGAGCGCGAAGGTATAACGTATAAATCATAGTAAACAACTCAGAAATG 19542
QY 49 -----GlyArgTyrIleProLeu-LeuAspTyrValGlyTyrAsnAsnTyrGlyTyrAs 66
Db 19543 GACCCGAGCGCTGGTCCCGCTAGCTCTCCAGCTCTCCCTGGCCAGCCCGAAGGAGA 19602
QY 66 nGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTyrAlaAlaValG1 86
Db 19603 GGGTCCGCATCCCTCCCGGTTCTCTCTC----- 19633
QY 86 yMetTyrThrAsnSerTyrAspProAsnAspGlyAla-----IleLeuArgSerSe 103
Db 19634 -----CTGGGTACCTGGCCTTGAGGTGGGGAAACGAGCCTACTTCTGTACCGTCTTT 19686
QY 103 rAspGlnGlyAlaThr-----TyrGlnI1 111
Db 19687 TGCCGACGGCGGACCCAGTGAATTAGCCGTTGGAGCCGCGAGCCTGCCTGGCTTTG 19746
QY 111 eThrProLeuProPheLysLeuGlyGlyAsnMetProGlyValArgGlyMetGly----- 128
Db 19747 CGCACCGAGTCTTGG-----GGACCTGGTGTCTCCCGGAAAACTTGGGAGCTGTGTA 19800
QY 129 -----GluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyr----- 142
Db 19801 TCCCGGGAGAGGCTTGGGACCTGGTGTCCCGGAGAGGCTTGGTACCTGGTTTCTCT 19860
QY 143 -----PheGlyAlaProSerGlyLysGlyLeu----- 151
Db 19861 GGAAGAGCCTTGGACACCTGTGTCTCGGAGGCGCTTGGGACCTGTGTCTCTGGGAGA 19920
QY 152 ---TyrArgSerThrAspSerGlyAlaThrTyrSerClnMetThrAsnPheProAspVa 170
Db 19921 GGCTTGGAGATCTGTGTCTCGGAGAGGCTTGGGGA----- 19957
QY 170 lGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyVa 190
Db 19958 -----CCTGGTGTCCCTGGAGA 19974
QY 190 lValTyrValAlaPheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIlePh 210
Db 19975 GGCTTGG-----GGACCTGGTGTACCTTGGAGAGGCTTGGAGACCTGGTGT 20019
QY 210 eValGlyValAlaAspProAsnAsnPro-ValPheTyrSerArg----- 224
Db 20020 TCTGGGAGAGGCT-----TGGGGACCTGGTGTCTTGGGAGAGGCTTGGGACCTGGTGT 20073
QY 225 -----AspGlyGlyAlaThrTyr-----GlnAlaValProG 235
Db 20074 CTCTGGAAAGAGGCTTGGACACCTGTGTACCCGGAGGCGCTTGGGATCTGGTGTCCCG 20133
QY 235 lyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeut 255
Db 20134 GAGAGCCTTGGGA-----CCTGGTGTCTCTG----- 20159
QY 255 yrlleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTyrLysP 275
Db 20160 -----GGAGAGGCTTGGGACCT-----GGTGACCTTGGAGAGGCTTGGGAC 20202
QY 275 heSerValThrSerGlyThrTyrArgIleSerProValProSerThrAspThrAlaA 295
Db 20203 CTGGTGTCTGAGAGAGCCTTGGGA-----TCTGGTGTCCAGAGAGGCTTGGGAC 20256
QY 295 snAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetV 315
Db 20257 CT-----GGTCTCTCGAGAGGCTTGGAC----- 20282
```

---

```
QY 315 alAlaThrGlnIleSerTyrTrpPro-----AspThrIleIlePheArgS 330
Db 20283 -----ACCTGGTGTCTCTGGGAGAGGCTTGGGACCTGGTG----- 20318
QY 330 erThrAspGlyGlyAlaThrTyrArgIleTyrAsp-TyrThrSerTyrProAsnArg 349
Db 20319 -----TCCTGGGAGAGGCTTGGGAGCCTGGTGT-----CCTGGGAGA 20355
QY 350 SerLeu-ArgTyrValLeuAspIleSerAlaGluProTyrLeuThrPheGlyValGlnPr 369
Db 20356 GGCTTGGAGA-----TCTGGTGAGCCGGAGAGGCTTGGGGA----- 20392
QY 369 oAsnProProValProSerProLysLeuGlyTyrMetAspGluAlaMetAlaIleAspPr 389
Db 20393 ---CCTGGTGTCCCGGAGAGGCTTGGGACTTGGTGTCTCCGGAGAGGCTTGGACAC 20448
QY 399 o-PheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspL 409
Db 20449 TGGTGTCCAGAGAGGCTT-----GGGACCTGGTGTACCTTGGAGAGGCTTGGGAC 20502
QY 409 euThrLysTyrAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluG 429
Db 20503 TGGTGACCCGGAGAGCCTTGGG----- 20525
QY 429 luThrAlaValAsnAspLeuIleSer-----ProProSerGlyAlaProLeuIleSerA 447
Db 20526 -----GACCTGTGTCTCTGGGAGAGCCTTGGGACCT----- 20558
QY 447 laLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSerThrIleP 467
Db 20559 ---GGTGACCTTGGAGAGGCTTGGGACCTGGTGTCTCTCGGAGAGTGCCTTGGGACCT- 20612
QY 467 heThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerI 487
Db 20613 ---AGTGACCCGGAGAGGCTTGGGACCTGGTGTCTCCGGAGAGGCTTGGGACCTGGTG 20670
QY 487 leIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHisValAlaP 507
Db 20671 TCCTG-----GGAGAGCCTTGGGATCTGGTGTCTCTG- 20702
QY 507 heSerThrAspGlyGlyLysAsnTyrPheGlnGlySerGlyProGlyGlyValThrThrG 527
Db 20703 ---GGGAGAGGCTTGGGACCTGTGTCTCTCGGAGAGAGCCTTGGGACCTTGTGAC 20757
QY 527 lyGlyThrVal---AlaAlaSerAlaAspGlySerArgPheValTyrAlaProGlyAspP 546
Db 20758 CGGAGAGGCTTGGACACCTGTGTCTCCGGGAGAG-----GCTTGGGACCTTGGTGAC 20811
QY 546 roGlyGln-----ProValValTyrAlaValGlyPheGlyAsnSerTyrAlaAlaS 563
Db 20812 CGGAGAGCCTTGGGACCTGTGTCTCTGGGAGAGGCTTGGGG-----ACCTGGTGTCTG 20868
QY 563 erGln-----GlyValProAlaAsnAlaGlnIleArgSerAspArgValA 578
Db 20869 GGAGAGAGCCTTGGGACCTGTGTGACCCGGAGAGGCTTGG----- 20909
QY 578 snProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGly----- 596
Db 20910 -----ACCTGGTGTCTG 20922
QY 597 --GlyValThrPheGlnProValAlaAlaGlyLeuPro---SerSerGlyValAlaGlyV 615
Db 20923 CGGAGAGCCTTGGGAGCCTGTGTCTCCGGGAGAGCCTTGGGAGCAGGTGACCTTGGAG 20982
QY 615 alMetPheHisAlaValProGlyLysGluGlyAspLeuTyrLeuAlaIaSerSerGlyL 635
Db 20983 AGGCTTGGGA-----CCTGTGTATCTTGGAGAGGCTTGGGACCTTGTGTCTCGGA- 21035
QY 635 euTyrHisSerThrAsnGlyGlySerSerTyrPheAlaIleThrGlyValSerSerAlaV 655
Db 21036 -----GAGTTACGGGGCTGTGTGGGGAGAGACGTTGTGTGAGCCAAAGTCCC 21084
QY 655 alAsn----- 656
```







```
QY 427 uGIuGlThrAla---ValaAspLeuIleSerProProSerGlyAlaProLeuIleSe 446
Db 3363 CGATGCCAGCGCAACGCTGTGTGACGTCGTACGCTAGCGGCTCCGCGAGC-----GA 3416

QY 446 rAlaLeuGlyAsp-----LeuGlyGly-----Ph 454
Db 3417 CGCCGAGGTGACACGCTGACGTACCACTGGGACGATCGTGGCAGGATGTGACCT 3476

QY 454 eThrHiAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrTh 474
Db 3477 GACGGGCGCCGAC---ACCCTGACCGCTCG-----TTACCGCGCGGACGACGCGCAG 3527

QY 474 xGlyThrSerValAspTyrAlaGluLeuAsnProSer----- 486
Db 3528 CGGTACCAACGCTCGGCTCATC---CTGACGGTCAGCGATGGCACCTCGACACACGAGGA 3584

QY 487 -----IleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspAr 503
Db 3585 TGCCGTGCGGCTCATCGTGGCGG-----GACCGGGTGCCACCCGAGC----- 3630

QY 503 gHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGl 523
Db 3631 -AACACTGCTCCGGAAGCGACGCGCGCGAGTCGCGCATCGTG---GCGAGGCGGCCAC 3686

QY 523 yValThrGlyGlyThrValAlaAlaSerAlaAspGlySerArg-----PheValTr 541
Db 3687 TGTACGCTCAACGGCAGC---GCGACGACGCTGACGCGTGACGCTGCTCATCGCTGTG 3743

QY 541 pAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAl 561
Db 3744 GACGAG---ATCGGCGGACGCGCGTG-----ACGTGTCTGACGGCTCTCGCTGAC 3794

QY 561 aAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAlaAsnProLysTh 581
Db 3795 GCCGAGCTTACCGCGCGCGCTCTCTCC-----GATCGGCTGAC 3833

QY 581 rPheTyrAlaLeu---SerAsnGly-----ThrPheTyrArgSerThrAspGlyGlyVa 598
Db 3834 CTTCCTCTCGTGTGTCAGCAGCGTACGCGGACACTGCTCGACGTCGACGTCATCAGCGT 3893

QY 598 lThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHi 618
Db 3894 GACCGAGAGAACGTGCGG-----CCGGTGGCGTCCGCTCGCGCGGTGCTC----- 3939

QY 618 sAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisE 638
Db 3940 -----TCTGGCAACACGACGTCGCGCACGCTCGACGCTCGGCTTCCAG 3983

QY 638 rThrAsnGlyGlySer-----SerTrpSerAlaIleThrGlyValSerSerAl 654
Db 3984 CGACGCGAACGCGATGTGCTGACGTACCGCTGGACGAGGTCTCCGCGCCGGAACGCGAC 4043

QY 654 aValAsn 656
Db 4044 CATCTCC 4050
```

## RESULT 14

```
US-09-902-540-1119
; Sequence 1119, Application US/0902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIORITY APPLICATION NUMBER: 60/217,883
; PRIORITY FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
```

```
; SEQ ID NO 1119
; LENGTH: 16584
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-1119

Alignment Scores:
Pred. No.: 6,65e-05 Length: 16584
Score: 175.50 Matches: 165
Percent Similarity: 35.27% Conservative: 90
Best Local Similarity: 22.82% Mismatches: 240
Query Match: 4.35% Indels: 229
DB: 4 Gaps: 47

US-09-917-376-3 (1-740) x US-09-902-540-1119 (1-16584)

QY 24 PheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArg 43
Db 6448 TACAACCTCAGTTCCCGGCGATGCTCCTCGACCATCGACTGGGCACGGCTACGCC 6507

QY 44 TrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAsp-----TrpValGly 59
Db 6508 GCGAAGCGGTGACGTCCTCGGTTCGGGTGCGTACGAGACGCGCGCTACACGGC 6567

QY 60 Trp-----AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsn 77
Db 6568 TGGTGTCTGGACGACCTCGAGTTCAACGGCATC----- 6600

QY 78 ThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGly 97
Db 6601 ACGAACACTCGGTCGCCACCATCGC-----CCCGAGGATGGC 6639

QY 98 AlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLys 117
Db 6640 GTTGT-----GTGAACCCCTGGCCCATCGC 6666

QY 118 LeuGlyGlyAsnMetProGlyArgGlyMet-----GlyGluArgLeu----- 131
Db 6667 AACGCGGT-----CCGACCGGTCCATCGCGCGGGTGAAGTGTGAGCCCTACGGT 6720

QY 132 ---AlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGly 150
Db 6721 TCGGCTCGCATCGGAGGCGCGCGTGCACCTTC----- 6756

QY 151 LeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspVal 170
Db 6757 -----ACCTGGGCACAG---ACGTCTGGCTCGCGGTG 6786

QY 171 -----GlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIle 187
Db 6787 ACCCTGGCGGGTCCCAACACCTGAACCGCTGTTACCGCGCGCGGCGGTTACCGAGTCC 6846

QY 188 GlnGlyValTrpValAlaPheAspLysSerSerSerSerLeuGlyGlnAlaSerLys 207
Db 6847 ACCGCACTGGTGTTCACGCTG-----ACCGTCTCGACGCGCTCAAGACGTCCACGAG 6900

QY 208 ThrIlePheValGlyValAlaAspProAsnProValPheTrpSerArgAspGlyGly 227
Db 6901 TCGGTGACCGGTGACGCTGCGCTCCCGCAACATCCG----- 6936

QY 228 AlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHisLysGlyValPhe 247
Db 6937 -----CCACCGTGAACCGGGGCTCGACGGCATCGTCGAGAGCGCGGTGAGTAC 6987

QY 248 AspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGly 267
Db 6988 -----ACGCTGAGCGGCTCCGCCACGATGCGGATGGC 7020

QY 268 SerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArgIleSerPro 287
Db 7021 AATGCC-----CTCACGTACCTGTGGAGCCCGAGGTCTCCGGT 7056

QY 288 ValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArg 307
```



Db 7057 ACGCCGGT-----GCGGTGAAGGACTAC----- 7080  
Qy 308 GlnHisProAnThrIleMetValAlaThrGlnIleSerTrpTrpProAspThrIle 327  
Db 7081 ACCACGCGCGGCGACGCTTCATCGCGCCGAGGTCACG---CTGGATGAGTCGCTGTC 7137  
Qy 328 PheArgSerThr-----AspGlyGlyAlaThrTrp-ThrArgIleTrpAspTrpTrpSe 345  
Db 7138 TTCTGCTCTGACGCTCAGCGACGGCATCGCCACGCTGAACG-----ACACG 7182  
Qy 345 rTyProAnArgSerLeuArgTyValLeuAspIleSerAlaGluProTrpLeuThrPh 365  
Db 7183 GTCACGCGTACGCTCACCACCGCA-----ACCGCGGCCCA----- 7219  
Qy 365 eGlyValGlnProAnProProValProSerProLysLeuGlyTrpMetAspGluAlaMe 385  
Db 7220 ----TCGCTCTCGACACGCTCGGCTTCGCGCGGCACT-GTCACCGTGACGCGCTC 7274  
Qy 385 tAlaIleAspProPheAnSerAspArgMetLeuTyrgly-----ThrGlyAlaTh 402  
Db 7275 CCGCGTGCACCGG---GATGGTACGCGCTACAGCTGGGAGCAGCGGGTGTTC 7331  
Qy 402 rLeuTyAlaThrAsn----- 407  
Db 7332 GAGGTGCGCCATCAATGGCGCGCACGCTCCGCCATCTCTTCGCCACCCCGTCCCGG 7391  
Qy 408 -----As 408  
Db 7392 TTCGTACGAGTTCCACGTGACGCGGACGCGCTCGGCTCTGCTCCACGCGGTGCC 7451  
Qy 408 pLeuThrIysTrpAspSerGlyGlnIleHisIleAlaProMetValLys---GlyLe 427  
Db 7452 GGTGACCATCATCGATGGTTCTCTCGCGCGCAACTCTGCGCGACCGGTGAACGCGGCAT 7511  
Qy 427 uGluGluThrAla---ValAsnAspLeuIleSerProProSerGlyAlaProLeuIleSe 446  
Db 7512 CGATGCCACGGGAACGCTGGTGACGTGCTGACGCTCAGCGCTCCGCGAGC-----GA 7565  
Qy 446 xAlaLeuGlyAsp-----LeuGlyGly-----Ph 454  
Db 7566 CGCGGAGGTGACACGCTGACGTACCATGCTGGGAGCAGATCGTGGCAGGATGTGACCT 7625  
Qy 454 eThiHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrTh 474  
Db 7626 GACGGCGCGCGAC---ACCCTGAGCGCTG-----TTACCGCGCGCGAGCAGCGCCAG 7676  
Qy 474 rGlyThrSerValAspTyAlaGluLeuAsnProSer----- 486  
Db 7677 CGGTACACGCTCGGCTTCATC---CTGACGCTCAGCGATGCGACCTCGACACCGACGCA 7733  
Qy 487 -----IleIleValArgAlaGlySerPheAspProSerSerGlnProAnAspAr 503  
Db 7734 TGCCGTGCGCGTCATCGTGGCGCG-----GACCGGGTGCCACCCGAGC----- 7779  
Qy 503 gHisValAlaPheSerThrAspGlyLysAsnTrpPheGlnGlySerGluProGlyGl 523  
Db 7780 -AACACTGCTCCGAGAGCGACGCGGCGAGTCGCGCATCGTG---GCGGAGGCGGCCAC 7835  
Qy 523 yValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArg-----PheValTr 541  
Db 7836 TGTGACGCTCAACGCGACG---GCGACCGACGCTGACGCTGACGCTGTCATCGTCTG 7892  
Qy 541 pAlaProGlyAspProGlyGlnProValThrAlaValGlyPheGlyAsnSerTrpAl 561  
Db 7893 GAGCGAG---ATCGGCGGCGACGCGCCGCGT-----ACGCTGTCTGACGCGCTCTCGCTGAC 7943  
Qy 561 aAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysTh 581  
Db 7944 GCGGAGTTTCACCGCCCGCGCTCTCTCC-----GATCGCTGAC 7982  
Qy 581 rPheTyAlaLeu---SerAsnGly-----ThrPheTyArgSerThrAspGlyGlyVa 598

7983 CTTCTCTCTGATGTGTCAGCGAGTACGGCGACTGTCGTCGACGTGACGTCCATCAGCGT 8042  
Qy 598 lThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheH 618  
Db 8043 GACCGAGAGAACGTCGCG-----CCGCTGGCGTCCGCTCGCGCGTGTCTC----- 8088  
Qy 618 sAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyHis 638  
Db 8089 -----TCTGGCAACACGACGCTCGCGCCACGCTCGACGCTCGCGCTTCCTCCAG 8132  
Qy 638 rThraSnglyGlySer-----SerTrpSerAlaIleThrGlyValSerSerAl 654  
Db 8133 CGACGCGAACGCGAGTGTCTGACGTACGCTGGACGCGAGGTCTCCGCGCCGACCGCAG 8192  
Qy 654 aValAsn 656  
Db 8193 CATCTCC 8199  
RESULT 15  
US-09-453-702B-39  
Sequence 39, Application US/09453702B  
Patent No. 6365723  
GENERAL INFORMATION:  
APPLICANT: Blattner, Frederick R.  
Perna, Nicole T.  
Plunkett, Guy  
Welch, Rod  
TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157  
NUMBER OF SEQUENCES: 265  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Quarles & Brady  
STREET: 1 South Pinckney Street  
CITY: Madison  
STATE: WI  
COUNTRY: US  
ZIP: 53701-2113  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 8.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/453,702B  
FILING DATE: 03-Dec-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/110,955  
FILING DATE: 04-DEC-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Seay, Nicholas J.  
REGISTRATION NUMBER: 27386  
REFERENCE/DOCKET NUMBER: 960296.95017  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (608) 251-5000  
TELEFAX: (608) 251-9166  
INFORMATION FOR SEQ ID NO: 39:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 25165  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
SEQUENCE DESCRIPTION: SEQ ID NO: 39:  
US-09-453-702B-39  
Alignment Scores:  
Pred. No.: 0.000375 Length: 25165  
Score: 170.50 Matches: 192  
Percent Similarity: 29.22% Conservative: 83  
Best Local Similarity: 20.40% Mismatches: 269  
Query Match: 4.22% Indels: 397  
Gaps: 48  
DB:







```

QY 578 -----AsnProLysThrPheTyrAlaLeuSer-----AsnGly 588
Db 20332 CACACCTATAACCTGAACAACATATTACACGCTGCTTCGTCATCAACACGGGGAAATGGG 20391
QY 589 ThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuPro 608
Db 20392 ACGTTTGTCTGGGGCAGAACACCAACCAATACCTTCCTGAGCGGGCGGCAGTGGCGCT 20451
QY 609 SerSerGlyAlaValGlyValMetPheHisAlaValProGlyLys----GluGlyAspLeu 627
Db 20452 ATGAGCAGCAGCGTTTCCATGACCTGGCGCGATTCGATGTGACGGCGATATGGATCTC 20511
QY 628 TrpLeuAlaAlaSerSerGly-----LeuTyrHisSerThrAsnGly 641
Db 20512 TTCTGCGCGCCAGCCAGGGAAGAGCTAACTACGGCTCGCTGTTATTCAACACCAACCGC 20571
QY 641 ----- 641
Db 20572 GTACTGGGTTGCCCGGTGGCGGCGCAACGCCAACCCTACGCCAGCCAGTTTAGC 20631
QY 642 -----Gly 642
Db 20632 CTGGCGGTGGACTGGAACCAACGACGGCCTGATGGATATGCGCGTATCGCCAGACCGGG 20691
QY 643 SerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSer 662
Db 20692 CAGTCGTAT---CTTTATACTAACGTCAGCAACGCCAGCAAC-----TGGACGCAATCG 20742
QY 663 AlaProGlySerSerTyrProAlaValPheValValGlyThrIleGlyGlyValThr--- 681
Db 20743 GCCCTCGCGCGCAGCCAGC-----GGTACCCAGCGCGGTGGCGCA 20787
QY 682 -----GlyValatyrArg-----SerAspAspCysGlyThrThr 692
Db 20788 ATGGACTACGACTGGGACGGCGGTGGTGGTGTCTCCAAACAGTCGGGCGACGCTG 20847
QY 693 TrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGly 712
Db 20848 TTCTGAGCCGCAACACCAACACCGGTGAGCTACGGC-----ACTTCG 20889
QY 713 AspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGly 732
Db 20890 CTACACCTGGCGATC-----ACCGATCCCAACGGCATTAACGTCTATTACGGC 20937
QY 733 Asp 733
Db 20938 AAT 20940

```

Search completed: October 5, 2005, 22:56:47  
 Job time : 5746.23 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 5, 2005, 08:30:02 ; Search time 1425.51 Seconds  
(without alignments)  
3584.301 Million cell updates/sec

Title: US-09-917-376-3  
Perfect score: 4036  
Sequence: 1 ATTQPTWSNVAIGGGFVD.....YIGTNGRGIVYDGGAPSG 740

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 7442561 seqs, 3452328358 residues

Total number of hits satisfying chosen parameters: 14885122

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US09917376/runat\_04102005\_164329\_26954/app.query.fasta\_1.2524  
-DB=Published Applications NA -QWMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR=SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=2000000000 -USRR=US09917376 @CNG 1 1 2178 @runat\_04102005\_164329\_26954  
-NCPU=6 -ICPU=3 -NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:  
3: /cgn2\_6/ptodata/1/pubpna/US05\_NEW\_PUB.seq:  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:  
13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:  
14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:  
15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:  
16: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:  
17: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:  
18: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq:  
19: /cgn2\_6/ptodata/1/pubpna/US10G\_PUBCOMB.seq:  
20: /cgn2\_6/ptodata/1/pubpna/US10H\_PUBCOMB.seq:  
21: /cgn2\_6/ptodata/1/pubpna/US10I\_PUBCOMB.seq:  
22: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:  
23: /cgn2\_6/ptodata/1/pubpna/US11A\_PUBCOMB.seq:  
24: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq:  
25: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:  
26: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	4036	100.0	2869	11	US-09-917-376-2	Sequence 2, Appli
2	4036	100.0	2869	15	US-10-155-400-2	Sequence 2, Appli
3	2429.5	60.2	2846	15	US-10-156-761-1845	Sequence 1845, Ap
4	2429.5	60.2	9025608	15	US-10-156-761-1	Sequence 1, Appli
5	1579	39.1	2517	15	US-10-026-994-4	Sequence 4, Appli
6	1579	39.1	2710	15	US-10-026-994-1	Sequence 1, Appli
7	1405	34.8	2217	15	US-10-156-761-2561	Sequence 2561, Ap
8	1405	34.8	9025608	15	US-10-156-761-1	Sequence 1, Appli
9	1139	28.2	3668	10	US-09-927-827-21	Sequence 21, Appli
10	1117	27.7	2367	18	US-10-395-241-13	Sequence 13, Appli
11	1117	27.7	2646	18	US-10-395-241-11	Sequence 11, Appli
12	1114	27.6	2481	18	US-10-395-241-17	Sequence 17, Appli
13	878.5	21.8	5698	18	US-10-420-191-1	Sequence 1, Appli
14	392.5	9.7	1103	20	US-10-653-047-7511	Sequence 7511, Ap
15	238	5.9	7407	16	US-10-246-330-3	Sequence 3, Appli
16	238	5.9	7407	17	US-10-282-122A-30151	Sequence 30151, A
17	199.5	4.9	7305	17	US-10-282-122A-11269	Sequence 11269, A
18	195	4.8	5295	17	US-10-282-122A-11871	Sequence 11871, A
19	192	4.8	9903	17	US-10-282-122A-28185	Sequence 28185, A
20	188.5	4.7	4399	17	US-10-369-493-37642	Sequence 37642, A
21	184	4.6	3324	17	US-10-282-122A-26289	Sequence 26289, App
22	182	4.5	69350	18	US-10-041-018-379	Sequence 379, App
23	180	4.5	2355	17	US-10-369-493-37792	Sequence 37792, A
24	179.5	4.4	3222	17	US-10-282-122A-14432	Sequence 14432, A
25	176	4.4	32367	14	US-10-158-160A-14	Sequence 14, Appli
26	174	4.3	2855	24	US-11-097-143-40748	Sequence 40748, A
27	174	4.3	5221	24	US-11-097-143-26683	Sequence 26683, A
28	174	4.3	5302	24	US-11-097-143-40747	Sequence 40747, A
29	173	4.3	2708	24	US-11-097-143-26684	Sequence 26684, A
30	170.5	4.2	2176	24	US-11-097-143-14408	Sequence 14408, A
31	170.5	4.2	4352	24	US-11-097-143-14407	Sequence 14407, A
32	170.5	4.2	7191	17	US-10-282-122A-35048	Sequence 35048, A
33	170.5	4.2	25165	14	US-10-114-170-39	Sequence 39, Appli
34	170	4.2	4323	17	US-10-282-122A-8383	Sequence 8383, Ap
35	169.5	4.2	3360	15	US-10-245-802-11	Sequence 11, Appli
36	169	4.2	69350	18	US-10-041-018-379	Sequence 379, App
37	168	4.2	5451	15	US-10-101-510-597	Sequence 597, App
38	167.5	4.2	2110	13	US-10-027-632-97984	Sequence 97984, A
39	167.5	4.2	2110	17	US-10-027-632-97984	Sequence 97984, A
40	167	4.1	1839	17	US-10-369-493-34868	Sequence 34868, A
41	166	4.1	3879	17	US-10-282-122A-8717	Sequence 8717, Ap
42	165.5	4.1	4393	15	US-10-245-802-21	Sequence 21, Appli
43	165	4.1	6615	17	US-10-282-122A-28180	Sequence 28180, A
44	165	4.1	67167	21	US-10-937-730A-3	Sequence 3, Appli
45	164.5	4.1	2540	22	US-10-680-860A-161	Sequence 161, App

ALIGNMENTS

RESULT 1  
US-09-917-376-2  
; Sequence 2, Application US/09917376  
; Publication No. US20040038334A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: 40197.4US01  
; CURRENT APPLICATION NUMBER: US/09/917,376  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1







Db 1999 CCTGGAAAGAGCGCATCTGTGGCTCGCTGTCATCGAGCGGCGCTTTACCACTCAACCAAT 2058  
Qy 641 GlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGly 660  
Db 2059 GCGCGACAGTGTGTCATCAACCGCGGTATCCTCCGCGGTGAACGTGGGATTTGGT 2118  
Qy 661 LysSerAlaProGlySerSerTyProAlaValPheValValGlyThrIleGlyVal 680  
Db 2119 AGTCCTGCGCGCGGTGTCATACCGACCGCTTTGTGTCGCGCACGATCGGAGCGTT 2178  
Qy 681 ThrGlyAlaTyArgSerAspCysGlyThrThrTrpValLeuIleAsnAspGln 700  
Db 2179 ACGGGGCGTACCGCTCCGACGACTGTGGACGACCTGGGTACTGATCAATGATACCCAG 2238  
Qy 701 HisGlnTyGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgVal 720  
Db 2239 CACCAATACGGAATTTGGGGAACAGCAATACCGGTGACCGCGAATTTACGGCGGGTG 2298  
Qy 721 TyrIleGlyThrAsnGlyArgGlyIleValTyGlyAspIleGlyGlyAlaProSerGly 740  
Db 2299 TACATAGGACGACGCGCGTGGAAATTTGATATACGGGGAATTTGGTGGCGCTCCGA 2358

## RESULT 2

US-10-155-400-2  
; Sequence 2, Application US/10155400  
; Publication No. US20030108988A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT AMYLASE FROM ACIDOTHERMUS  
; FILE REFERENCE: CELLULOYTICUS  
; CURRENT APPLICATION NUMBER: US/10/155,400  
; CURRENT FILING DATE: 2002-10-22  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2869  
; TYPE: DNA  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; NAME/KEY: modified\_base  
; LOCATION: (2869)  
; OTHER INFORMATION: a, c, t, g, other or unknown  
US-10-155-400-2

Alignment Scores:  
Pred. No.: 0 Length: 2869  
Score: 4036.00 Matches: 740  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 15 Gaps: 0

US-09-917-376-3 (1-740) x US-10-155-400-2 (1-2869)

Qy 1 AlaThrThrGlnProTyThrTrpSerAsnValAlaIleGlyGlyValAsp 20  
Db 139 GCGACGACTCAGCCGTACACCTGGAGCAACGTGGCGATCGGGGCGCGGCTTTGTCGAC 198  
Qy 21 GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyValArgThrAspIleGlyGly 40  
Db 199 GGGATCGTCTTCATGTGAGGTGACCGCGGAATTCGTGTCGCGGACGACATCGGGGG 258  
Qy 41 MetTyArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrp 60  
Db 259 ATGTATCGATGGATGCCGCAACGGGCGTGGATCCCTCTTCTGGATTTGGGTGGGATG 318  
Qy 61 AsnAsnTrpGlyTyArgGlyValValSerIleAlaAspProIleAsnThrAsnLys 80  
Db 319 AACAAATTTGGGGGTACAAACGCGCTGTCAGCATTTGGCGGACACCCGATCAATAACAAG 378

Qy 81 ValTrpAlaAlaValGlyMetTyThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 100  
Db 379 GTATGGGCGCGCTCGGAATGTACCAACAGCTGGGACCCAAACGACGCGAGCGATTCTC 438  
Qy 101 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGly 120  
Db 439 CGCTCGTCTGATCAGGGCGCAACGTGGCAATAAACCCTTCGCTTCAAGCTTGGCGGC 498  
Qy 121 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 140  
Db 499 AACATGCCCGCGCGTGGAAATGGCGGAGCGCTTTCGGTGGATCCAAACAATGACAAT 558  
Qy 141 LeuTyPheGlyAlaProSerGlyLysGlyLeuTyArgSerThrAspSerGlyAlaThr 160  
Db 559 CTGTATTTTCGGCGCCCGCGAGCGCAAGGGCTCTGGAGAAGCACAGATTCCGCGCGACC 618  
Qy 161 TrpSerGlnMetThrAsnPheProAspValGlyTyThrIleAlaAsnProThrAspThr 180  
Db 619 TGGTCCCAGATGACGAATTTCCGGACGTAGGCACGTACATTGCAAAATCCCACTGACAC 678  
Qy 181 ThrGlyTyArgGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSer 200  
Db 679 ACCGCTATCAGACGATATTCAAGCGCTCGTCTGGGTGCTTTCGACAAAGTCTTCGTCA 738  
Qy 201 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal 220  
Db 739 TCGCTCGGCAAGCGAGTAGACCATTTTGTGGGCTGGCGGATCCCAATAATCCGCTC 798  
Qy 221 PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 240  
Db 799 TTCTCGAGCAGAGACGGCGCGCGACGTGGCAGCGGTGCGGGTGGCGGACCGGCTTC 858  
Qy 241 IleProHisLysGlyValPheAspProValAsnHisValLeuTyThrIleAlaThrSerAsn 260  
Db 859 ATCCCGCAACAGGGCGCTTTTGACCCCGTCAACACGTGCTCTATATTGCCACGACGAA 918  
Qy 261 ThrGlyGlyProTyArgSerGlySerValTrpLysPheSerValThrSerGly 280  
Db 919 ACGGGTGGTCCGTATGACGGGAGCTCCGGGAGCTCTGGAAATTTCTCGGTGACCTCGG 978  
Qy 281 ThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyPheGlyTy 300  
Db 979 ACATGGACGCGAATCAGCCCGGTACCTTCGACGACACGCGCGGCTGCGGACGCGGATC 1038  
Qy 301 SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer 320  
Db 1039 AGCGGCTCTACTATCGACCGCCAGCACCCGAAACACGATATATGGTGGCAACCCAGATAT 1098  
Qy 321 TrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIle 340  
Db 1099 TGGTGGCGGACACCATTAATCTTTTCGAGACCGCGCGGCTGGAGCTGGACGCGGATC 1158  
Qy 341 TrpAspTrpThrSerTyProAsnArgSerLeuArgTyValLeuAspIleSerAlaGlu 360  
Db 1159 TGGGATGGACGAGTTATCCCAATCGAAGCTTGGGATATGCTGTGACATTTGGCGGAG 1218  
Qy 361 ProTrpLeuThrPheGlyValGlnProAsnProValProSerProLysLeuGlyTrp 380  
Db 1219 CCTTGGCTGACCTTCGGCGTACAGCCGAATCTCTCCGTTACCCAGTCCGAAAGCTCGGCT 1278  
Qy 381 MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyThrGly 400  
Db 1279 ATGGATGAACGCGATGCAATTCGATCCGTTCAACTCTGTATCGAAGTCTTACGGAACAGC 1338  
Qy 401 AlaThrLeuTyAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIle 420  
Db 1339 GGGACCTTGTACCCAAACAATGATCTCACGAGTGGGACTTCGCGCGGCGCAGATTCAATC 1398  
Qy 421 AlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProProSer 440  
Db 1399 GCGCCGATGGTCAAGAGGATTTGGAGGACGCGGTAACGATCTCTCATCAGCCCGCTCT 1458



```
QY 441 GlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThr 460
Db 1459 GCGCGCGCTCATCAGCGCTCTCGAGACCTCGCGGCTTCAACCAGCGCGAGTTACT 1518
QY 461 AlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyr 480
Db 1519 GCGGTGCATCGACGATCTTCACGTCAACGGTGTTCACGACCGGACCGAGCGTGCATAT 1578
QY 481 AlaGluLeuAenProSerIleValArgAlaGlySerPheAspProSerSerGlnPro 500
Db 1579 GCGGAATGAATCCGTCGATCATGCTTCGCGCTGGAAGTTTCGATCCATCCAGCCAAACG 1638
QY 501 AsnAspArgHisValAlaPheSerThrAspGlyGlyLyAsnTrpPheGlnGlySerGlu 520
Db 1639 AACGACGACGCTCGCGTCTCGACACGCGCGCAAGAACTGTTTCCAAAGGCGAGAA 1698
QY 521 ProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheVal 540
Db 1699 CCGCGCGGTGACGACGCGGCGCACCGCTCGCGCATCGGCGGACGGCTCTCGTTTCGTC 1758
QY 541 TrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrp 560
Db 1759 TGGGCTCCCGCGCATCCCGGTGAGCTGTGTGTACGCGAGTCGGATTTGGCAACTCCCTGG 1818
QY 561 AlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLys 580
Db 1819 GCTGCTTCGCAAGGTGTTCCCGCAATGCCAGATCCGCTCAGACCGGCTGAATCCAAAG 1878
QY 581 ThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPhe 600
Db 1879 ACTTCTATGCCCTATCCAAATGGAACCTTCTATCGAAGCAGCGCGCGGTGACATTC 1938
QY 601 GlnProValAlaAlaGlyLeuProSerSerGlyAlaValAlaGlyValMetPheHisAlaVal 620
Db 1939 CAACCGGTCGCGCGCGCTTCCTCGAGCAGCGGTGCGCGTGTCAATGTTCCACGCGGTG 1998
QY 621 ProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsn 640
Db 1999 CCTGAAAGAAGGCGATCTGTGGCTCGCTGATCGACGCGGCTTTACCATCAACCAAT 2058
QY 641 GlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGly 660
Db 2059 GCGCGCAGCAGTTGTCTGCAATCACCGCGGTATCTCTCGCGGTGAACGTGGGATTTGGT 2118
QY 661 LysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGlyVal 680
Db 2119 AAGCTGCGCGCGGTGCTGATACACGCGCTTCTTGTGTCGCGCACCATCGAGGCGGT 2178
QY 681 ThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAspAspGln 700
Db 2179 ACGGGGCGGTACCGCTCCGACGACTGTGGAGCAGCTGGGTACTGATCAATGATGACCCAG 2238
QY 701 HisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArgVal 720
Db 2239 CACCAATACGGAAATTTGGGCAACAGCAATCACCGGTGACCGGAAATTTACGGCGGTG 2298
QY 721 TyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyGlyAlaProSerGly 740
Db 2299 TACATAGGACGAAACGCGCGTGGAAATTTGATACGCGGACATTTGTGTGTGCGCGTCCGGA 2358
```

## RESULT 3

US-10-156-761-1845

; Sequence 1845, Application US/10156761

; Publication No. US20030119018A1

; GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI

; APPLICANT: IKEDA, HARUO

; APPLICANT: ISHIKAWA, JUN

; APPLICANT: HORIKAWA, HIROSHI

; APPLICANT: SHIBA, TADAYOSHI

; APPLICANT: SAKAKI, YOSHIYUKI

; APPLICANT: HATTORI, MASAHIRA

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

```
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1845
; LENGTH: 2646
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2646)
US-10-156-761-1845
```

## Alignment Scores:

```
Pred. No.: 4,45e-224 Length: 2646
Score: 2429.50 Matches: 438
Percent Similarity: 73.9% Conservative: 111
Best Local Similarity: 59.0% Mismatches: 174
Query Match: 60.2% Indels: 19
DB: 15 Gaps: 8
```

US-09-917-376-3 (1-740) x US-10-156-761-1845 (1-2646)

```
QY 6 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 25
Db 115 TACAGCTGGAAGAACCCCGCGTGCAGCGCGCGCTTCGTCGCCGATCGTCTTCAAC 174
QY 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAsp 45
Db 175 CGTCCGAGAGAACCTCGCTACCGCCGACCGACATCGCGCGCGCTACCGCTGGGCC 234
QY 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr 65
Db 235 GAGTCTCGAAGACCTGGACGCGCTGCTCGACTCGTCCGCTGGAGCGACTGGGGGCAC 294
QY 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnIleValTrpAlaAlaVal 85
Db 295 ACGGTGTGTCGAGCGCTCGCTCCGACTCCGCGACCGCGAACAGAGGTGACGCGCGCTC 354
QY 86 GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGln 105
Db 355 GGCAGTACAGAACAGCTGGGACCGCGGCAACGCTGCTGCTCAGTCCGCGCGACCGG 414
QY 106 GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArg 125
Db 415 GCGCGAGCTGGCAGAGACCGACCTGCCCTTCAAGCTGGCGGGAACATGCCGCGCGG 474
QY 126 GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 145
Db 475 GGCATGGCGAGCGCTCGCGTGCACCCGAAACAGAACAGCGTGTGTATTCGCGCGG 534
QY 146 ProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 165
Db 535 CCCAGCGGCAAGGGCTGTGGCGGTGCGAGCTCGGGGCGCTCTCGTTCGACAGTCAAC 594
QY 166 AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSer 185
Db 595 GACTTCCGACGCTCGGACCTACGTGCGAGGACCGGACCGACAGCGGGTACCGGTCC 654
QY 186 AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeuGlyGlnAla 205
Db 655 GACAACAGGGCATCGTGTGGGTCACTTCGACGAGTCGACCGGGTTCGCGGGGAGCTCC 714
QY 206 SerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArgAsp 225
Db 715 ACGCGAGCGGTGTACGTGCGGGTCCGCGGACAGGACAACTCCGTCTCTCTCCACGGAC 774
QY 226 GlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHisLysGly 245
```



775 GCAGGCGGACCTGGTCCCGGCTGGCGGCGACCGCGCATCTCGCCCAAGGCG 834  
246 ValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyr 265  
835 GTGCTGGACGCGGCGAAGCGCTGTCTGTACCTCGGTACAGGACAAAGGCGCGACCGTAC 894  
266 AspGlySerSerGlyAspValTyrIlePheSerValThrSerGlyThrThrThrArgIle 285  
895 GACGGCGGACGAGGACGCTGTGGGGGTACACGACGAGACCGGACCTGGAGCAACATC 954  
286 SerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIle 305  
955 AGCCCGGTGGCGGAGCGGACAC-----TACTACGGCTTCAGCGGGCTGACCGGTG 1005  
306 AspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerThrThrThrProAspThr 325  
1006 GACCGCGACATCCGGGAGCGGTGTGGCGACTGTGCGTACAGCTTCCTGGTGGCGGACACG 1065  
326 IleIlePheArgSerThrAspGlyAlaThrThrArgIleTyrAspThrThrThr 345  
1066 CAGCTTTCCTCCACGACAGCGCGGACCTGGAGAGGCTGGGACTACACTCG 1125  
346 TyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProThrLeuThrPhe 365  
1126 TATCCGAGCGCTCGAACCGCTTCCACATGATGTCTCTCGTCCGCTCGCTCACCTGG 1185  
366 GlyValGlnProAsnProValProSerProTyrLeuGlyTyrMetAspGluAlaMet 385  
1186 GGAGGCGACCCGCGCGCGGACAGACCCGAACTCGGCTGGATGACCGAGTCCCTG 1245  
386 AlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAla 405  
1246 GAGATGACCGCTTCGAGTCCCGCGCATGATACGGAACCGGCGCGGCTTCACGGC 1305  
406 ThrAsnAspLeuThrLysThrAspSerGlyGlyGlnIleHisIleAlaProMetValLys 425  
1306 ACGGACACTGACGACTGGGACGACGAGGAGCAGTTCACCATCAAGCGGATGGCGCG 1365  
426 GlyLeuGluGluThrAlaValAsnAspLeuIleSerProProSer---GlyAlaProLeu 444  
1366 GGCTGGAGGAGAGCGGCGCTCAACGACCTCGCTCGCTCCCTCGCGCGCGCGCCAGCTG 1425  
445 IleSerAlaLeuGlyAspLeuGlyPheThrHisAlaAspValThrAlaValProSer 464  
1426 TTCAGCGCGCTCGTACATCGCGGCTTCGCGCACACGACGACCTCACCGGTGGCGTGC 1485  
465 ThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsn 484  
1486 CTGATGTACAGTTCGCGGCTTCGCGCACTTCGCGCACGACGACGCTCGACTACGCGGAC 1545  
485 ProSerIleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHis 504  
1546 CCGGCGACGCTGGTGGGCTGGCGCACTTCGAC-----TCGGGTCCG-----CAT 1590  
505 ValAlaPheSerThrAspGlyGlyLysAsnThrPheGlnGlySerGluProGlyVal 524  
1591 GTGGCGTTCGACGACGACACCGCGCACTGGTTTCGCGGGGCGGACCTTCGCGGGTGC 1650  
525 ThrThrGlyThrValAlaAlaSerAlaAspGlySerArgPheValTyrAlaProGly 544  
1651 AGCGGGGGTGGGACGCTGGCGCGCTCGAGCGGAGTCTGCTGTGTGGAGCCCGCGG 1710  
545 AspProGlnProValValTyrAlaValIleGlyPheGlyAsnSerThrPheAlaSerGln 564  
1711 GGACCGGG-----GTGCGAGTACACGACCGGGTTCGACCTCTGTCGGGTTCGCG 1764  
565 GlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAla 584  
1765 GGCCTCCCGCGCGGCGATCGTGGAGTCCGAGTCCGACCGGGTTCGACCGAAGACTTCTAC 1824  
585 LeuSerAsnGlyThrPheTyrArgSerThrAspGlyValThrPheGlnProValAla 604  
1825 TTCAGTCCGCGAGGTTCATCGTTCGAGTTCGAGCGGCGGCGGACCTTCACGCGGTTCGCG 1884

605 Ala---GlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLys 623  
1885 GCCACGGGCTCCGAGCGCGACAGC-----GTGCGCTTCAAGCGCTGCCCGGCGACG 1938  
624 GluGlyAspLeuTyrIleAlaSerSer-----GlyLeuTyrHisSer 638  
1939 AAGGCGACATCTGGCTGGCGCGCGAGCGAGCGCGCGTACCGGGCTGTGGCACTCG 1998  
639 ThrAsnGlyGlySerSerThrSerAlaIleThrGlyValSerSerAlaValAsnValGly 658  
1999 ACGGACGGCGCGCGCTTCCACAGCTCGCCACCGTCCGACCGCGGCGGACCATCGGC 2058  
659 PheGlyLysSerAlaProGlySerSerTyrProAlaValPheValGlyThrIleGly 678  
2059 TTCGCGAGCGCGGACCGCGCTGTACACAGCGCTTACACAGCGCGGAGATCGGC 2118  
679 GlyValThrGlyAlaTyrArgSerAspCysGlyThrThrThrValLeuIleAsnAsp 698  
2119 GGTGTGGCGGATCTTCGGTTCGACCGACGAGCGGCGGAGCTGGACCCGCGTCAACGAC 2178  
699 AspGlnHisGlnTyrGlyAsnThrGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArg 718  
2179 GATGCCACACGAGTGGGGTTGGACGGCGCGCGATCACCGGTGACCCAGGGTCTACGGG 2238  
719 ArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyGlyAlaPro 738  
2239 CGGTGTATGTGTCGACGAGCGGCGCGGATGCTCTACGCGGACACCGCGGGCTCTTCG 2298  
739 SerGly 740  
2299 GACGGC 2304  
RESULT 4  
US-10-156-761-1  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; PRIOR FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1  
; LENGTH: 9025608  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4187715)  
; OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1  
Alignment Scores:  
Score: 1, 41e-219 Length: 9025608  
Percent Similarity: 2429.50 Matches: 438  
Best Local Similarity: 73.99% Conservative: 111  
Query Match: 59.03% Mismatches: 174  
DB: 60.20% Indels: 19  
Gaps: 8  
US-09-917-376-3 (1-740) x US-10-156-761-1 (1-9025608)



QY	6	TyrThrTrpSerAsnValalaileGlyGlyGlyPheValaspGlyIleValPheAsn	25
Db	2277508	TACAGTGGAGAACCGCCGGTGCAGCGCGCGCTTCGTCGCCGGCATCGTCTTCAAC	2277567
QY	26	GlulGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArgTrpAsp	45
Db	2277568	CGCTCCGAGAGAACCTCGCCTACGCCCGCACCGACATCGCGGGCGCCTACCGCTGGGCG	2277627
QY	46	AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr	65
Db	2277628	GAGTCTCGAAGACCTGGACCGCGTGTCTGACTCGGTGGGTGAGCGACTGGGGGCAC	2277687
QY	66	AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal	85
Db	2277688	ACGGGTGTCTGAGGCTCGCCTCCGACTCCGACTCCGACCCCGAACAAGGTGTACGGCGCGTC	2277747
QY	86	GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGln	105
Db	2277748	GGCAGTTACACGAACAGCTGGGACCGCGCAACGGTGCCTGCTCAGGTCCGGCGACCGG	2277807
QY	106	GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArg	125
Db	2277808	GGCGGAGCTGGCAGAAGACCGACCTGCGCTTCAAGCTGGCGCGGAACATGCGCGGCGCG	2277867
QY	126	GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla	145
Db	2277868	GGCATGGGCGACCGCTCGCGTGCAGCCGCAACAGAACAGAGCTGCTGTATCTCGCGCGG	2277927
QY	146	ProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr	165
Db	2277928	CCACGCGCAAGGGCTGTGGCGTCTGACGGACTCGGGGCGCTCTGTGTCGAGGTCAAC	2277987
QY	166	AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrGlyTyrGlnSer	185
Db	2277988	GACTTCCCGAAGCTGGCGCACTACGTGCAGGACGCGACCGACACAGCGGGGTACCGTCC	2278047
QY	186	AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeuGlyGlnAla	205
Db	2278048	GACACACAGGGCATCTGTGGGTGCACCTTCGACGAGTGCAGGGGTGCGCGGGAGCTCC	2278107
QY	206	SerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArgAsp	225
Db	2278108	ACGCGGACGGTGTACGTGGGGTTCGCGCAAGGACAACTCGTCTATCGCTCCACGGAC	2278167
QY	226	GlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHisLysGly	245
Db	2278168	GCGGCGCGACCTGTGTCGGGTGGCGCGCGCACCGCCATCTCGCCCAACAAGGCG	2278227
QY	246	ValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyr	265
Db	2278228	GTGTGGGACCGCGGACCGGTGTGTGTACCTTCGGGTACAGCGACAAAGGGCGGACGTAC	2278287
QY	266	AspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArgIle	285
Db	2278288	GACGCGGCAAGGGACAGTGTGGCGGTACACGACGAAGACGGGACCTGGACGAATC	2278347
QY	286	SerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIle	305
Db	2278348	AGCCCGGTGCGGAGCGCGACAC-----TACTACGGGTTCAGCGGGCTGACCGGTG	2278398
QY	306	AspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAspThr	325
Db	2278399	GACCGGACGATCCGGGACCGGTGTGTGTACCTTCGCGTACAGCTTCCTGGTGGCGGACACG	2278458
QY	326	IleIlePheArgSerThrAspGlyAlaThrTrpThrArgIleTrpAspTrpThrSer	345
Db	2278459	CAGCTCTTCCTTCACGACAGCGCGCACCTTGACGAAGGCCTGGGACATACCTCG	2278518
QY	346	TyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPhe	365
Db	2278519	TATCCGAGCGCTCGAAGCCGCTTACCATGGATGTCTCGTCTCGCCCTGGCTCACTCG	2278578
QY	366	GlyValGlnProAsnProValProSerProLysLeuGlyTrpMetAspGluAlaMet	385
Db	2278579	GGAGGAACCCCGCACCCCGGAGCAGACCCCGAAACTCGGTGGATGACCGAGTCCCTG	2278638
QY	386	AlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAla	405
Db	2278639	GAGATCGACCCGCTTCGACTCCCGCGCATGATGTACGAACCGGCGCGCTCTACGGC	2278698
QY	406	ThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLys	425
Db	2278699	ACGGACAACCTGACCACTGGGACAGCGAAGCCAGTTCACCATCAAGCCGATGGCGCGG	2278758
QY	426	GlyLeuGluGluThrAlaValAsnAspLeuIleSerProProSer---GlyAlaProLeu	444
Db	2278759	GGCCTGGAGGAGACGGCGCTCAACGACCTCGCTCCCTCCGCGCGCGCCAGCTG	2278818
QY	445	IleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSer	464
Db	2278819	TTACGCGCGCTCGGTGACATCGCGCGCTTCCGCGCACACGGAOCTCAACCGGTGCGGTG	2278878
QY	465	ThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsn	484
Db	2278879	CTGATGTACACGTCCCGCACTTCCACAGACACACAGCTCGACTACGCCGAGACCGAC	2278938
QY	485	ProSerIleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHis	504
Db	2278939	CCGGCACCGGTGTCGGGTCCGCAATCTCGAC-----TCGGGTCCG-----CAT	2278983
QY	505	ValAlaPheSerThrAspGlyLysAsnTrpPheGlnGlySerGluProGlyGlyVal	524
Db	2278984	GTGGGTGTCTCGACGACAAACCGCGCAACTGTGTTCGCGGGCGGACCCCTTCGGGGGT	2279043
QY	525	ThrThrGlyGlyThrValAlaAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGly	544
Db	2279044	ACGGGGGTGGACCGTCCGCGCGCTCCGCGCAGTCGCTTCGTGTGGAGCCCGCGCG	2279103
QY	545	AspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGln	564
Db	2279104	GGCACCGGG-----GTGCAGTACACGACCGGGTTCGGCACCTCGTGGTCCGCGTCCGCG	2279157
QY	565	GlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAla	584
Db	2279158	GGCCTCCCGCGCGCGCATCGTCGATCCGACCGGGTCGACCCGAGACTTCTTACGGC	2279217
QY	585	LeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAla	604
Db	2279218	TTCAAGTCGGCAGGTTCACGTTCAGTCGAGCGGGGGGACCTTACCGGCTCCGCG	2279277
QY	605	Ala---GlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLys	623
Db	2279278	GCCACGGGCTGCCGAGCGCGCGACAGC-----GTGGCTTCAAGGGCTGCCCGGACG	2279331
QY	624	GlulGlyAspLeuTrpLeuAlaAlaSerSer-----GlyLeuTyrHisSer	638
Db	2279332	AAGGGCGCATCTGGCTGGCGCGCGCGAGCGAGCGCGGTGTGGCTGTGGCACTCG	2279391
QY	639	ThrAsnGlyLysSerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGly	658
Db	2279392	ACGGACGGCGCGCGCTTCACCAAGCTCCGACCGCGCGCGGTGACCGCGGACCATCGGC	2279451
QY	659	PheGlyLysSerAlaProGlySerSerTyrProAlaValPheValGlyThrIleGly	678
Db	2279452	TTCCGCAAGCGCGCACCGCGCTCTACACAGCTCTACACCGCGGAGAGATCGGC	2279511
QY	679	GlyValThrGlyAlaTyrArgSerAspCysGlyThrThrTrpValLeuIleAsnAsp	698
Db	2279512	GGTGTGGCGGCATCTTCGGTCCGACCGCAAGGGCGCGAGTGGACCCCGCTCAACGAC	2279571
QY	699	AspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArg	718
Db	2279572	GATGCCACAGTGGGTGGACGGCGCGCGATCACCGGTGACCCCGGTCTACGGG	2279631
QY	719	ArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyAlaPro	738







```
QY 532 AlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValVal 551
Db 1645 TATTTCGGCCGACGCGACGATCTCTGTGTCGACCGCTCGTGGG-----GTG 1695
QY 552 TyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGln 571
Db 1696 CAGCGCTCGAGTTCACAGGACGTTTGGCTCCGCTCGAGCCTGCGCGGGCGCGTC 1755
QY 572 IleArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyr 591
Db 1756 ATCGCTCGGACAAAGACCAACAGCGCTCTCTACGCGCGCTCGGATCGACCTTTTAC 1815
QY 592 ArgSerThrAspGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGly 611
Db 1816 GTACAGGAGGACCGGACGAGCTTC-----ACGCGCGG-----CCCAAGCTGGGC 1863
QY 612 AlaValGlyValMet-----PheHisAlaValProGlyLysGluGlyAspLeuTrpLeu 629
Db 1864 AGCGAGGAGATCCGGGATATCGTCTCACCGACCGCGGGGACGTTGTATGTC 1923
QY 630 AlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyLysSerTrpSerAlaAla 648
Db 1924 TCGACCGACGTCGCATATTCGCTCCACAGACTCGGGCAGACCTTTGGCCAAAGTCTCC 1983
QY 649 ThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyr 668
Db 1984 ACCCGCTGACCAACACTACAGATCGCCCTGGGTGTGGGCTCA---GGCTCGAAGCTGG 2040
QY 669 ProAlaValPheValValGlyThrIleGlyValThrGlyValA-----TyrArgSer 686
Db 2041 ---AACCTGTATGCTTCGGCACC-----GGCGGTGAGGGGTGCGCTCTAGCCAGT 2091
QY 687 AspAspCysGlyThrThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrp 706
Db 2092 GGAGACAGCGCGCTCTCTGACGACATCCAGGCTCCCGAGGGCTCGGCTCCATCGAC 2151
QY 707 GlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGly 726
Db 2152 AGCACCAAGTTCGCGCGCGCGACGACACCGCGGCAAGTCTACGTGGGACCAACGCG 2211
QY 727 ArgGlyIleValTyr-----GlyAspIleGlyValAlaProSerGly 740
Db 2212 CGGGCGGTCTTTTACGTACAGGAACCGTCCGCGCGGCGGCGG 2259

RESULT 6
US-10-026-994-1
; Sequence 1, Application US/10026994
; Publication No. US20030113732A1
; GENERAL INFORMATION:
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Goedegebuur, Frits
; APPLICANT: Ward, Michael
; APPLICANT: Yao, Jian
; TITLE OF INVENTION: EGV1 Endoglucanase and Nucleic Acids
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: GC698
; CURRENT APPLICATION NUMBER: US/10/026,994
; CURRENT FILING DATE: 2002-04-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2710
; TYPE: DNA
; ORGANISM: Trichoderma reesei
US-10-026-994-1

Alignment Scores:
Pred. No.: 6,27e-142 Length: 2710
Score: 1579.00 Matches: 325
Percent Similarity: 59.26% Conservative: 123
Best Local Similarity: 42.99% Mismatches: 264
Query Match: 39.12% Indels: 44
```

```
DB: 15 Gaps: 20
US-09-917-376-3 (1-740) x US-10-026-994-1 (1-2710)
QY 6 TyrThrTrpSerAsnValAlaIle---GlyGlyGlyGlyPheValAspGlyIleValPhe 24
Db 101 TTTTCATGGAAGAACGTCAGCTCGCGCGCGCGCTGCTGCCCGGATCATCTTC 160
QY 25 AsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrAlaGlyTrp 44
Db 161 CATCCCAAGACAAAGGCGTAGCATATGACGAAACAGATATTTGGCGGGCTGTACCGCCTC 220
QY 45 AspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpVal-----GlyTrp 60
Db 221 AAC---GCCGAGACTCATGACCGCGCTCACGGATGGGATGCTGATAATGCCGCTGG 277
QY 61 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 80
Db 278 CACAACCTGG-----GGCATCGACGCTGTGGCTTGATCCCGCAGCAGCATCAAAAG 328
QY 81 ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 100
Db 329 GTGTATGCCGCGAGTGGCATGTATACGAACAGCTGGGATCCGAGATATGGAGCATCATC 388
QY 101 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGly 120
Db 389 CGCTCGTCAGACCGCGCGCAACGCTGCTCTTCCCACTTGCCTTCAAAGTCGGGGT 448
QY 121 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 140
Db 449 AACATGCCAGGACGCGGAGCGCGAGAGCGTCTGGCTGTGATCCGCGCAACTCCAACATC 508
QY 141 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThr 160
Db 509 ATCTACTTTGGTGTCTGCTCAGGAAACCGGCTCTGGAAGTCTACGACGCGCGGTGACC 568
QY 161 TrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 180
Db 569 TTTTCCAAAGTCTCTGCTTACGGCAACTGGGACGTACATCCAGACCGAGTGATTC 628
QY 181 ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSer 200
Db 629 AACGCTGTACAACAGCGACCAAGCAAGGACTCATGTGGTTACGTTCTGACTCAACACGAGC 688
QY 201 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspPro---AsnAsnPro 219
Db 689 ACGACCGGGGAGCCACGCTCTGCTATCTTTGTGGCAGCGCTGATAAACAATCACTGCTTCA 748
QY 220 ValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGly 239
Db 749 GTCTATGTGAGCAGCAATGCCGCTCCAGCTGGAGTGTCTGTACCGGGGCGCAGGAGAAA 808
QY 240 PheIleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSer 259
Db 809 TACTTTCTCTACAAGCGCAAACTGCGACCCAGCAGAGAGGCGCTTGTATCTGACCTATTC 868
QY 260 AsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSer 279
Db 869 GATGCGACAGGCGCGCTATGATGGCACCTTGGCTCAGTGTGGAGTACGACATTTGACGGG 928
QY 280 GlyThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGly 299
Db 929 GGAACCTTGGAAAGACATCACCCCTGTCTCTGGATCAGATCTA-----TACTTTGGC 979
QY 300 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 319
Db 980 TTTGGCGGCTTGGCTCGATTTGCAAAAGCCAGGAAACCTTGTGTGTTGTTCTTCTTGAAC 1039
QY 320 SerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArg 339
Db 1040 TCTTGGTGGCCAGATGCTCAGCTGTTTTCGTCGACCGACTCTGGGACACATATGGAGCCG 1099
QY 340 IleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAla 359
```



Db 1100 ATCTGGCGTGGCGAGCTATCCGACTGAGACCTATTACTACGATCTCAACTCCCAA 1159  
Qy 360 GluProTrpLeuThrPheGly---ValGlnProAsnProProValProSer----- 375  
Db 1160 GCACCGTGGATCAAGAACAACTTTATCGATGTGACGAGCGAGTCCCGTGGTGTCTC 1219  
Qy 376 ---ProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg 394  
Db 1220 ATCAAGCCCTCGGCTGGATGATTGAGTCTCTCGAGATTGACCCAACGACGACCAACCAC 1279  
Qy 395 MetLeuTrpGlyThrGlyAlaThrLeuTrpAlaThrAsnAspLeuThrLysTrpAspSer 414  
Db 1280 TGGCTCTACGACCGCGGATGACATCTTGGCGGCCACGATCTCAACACTGGGACACG 1339  
Qy 415 GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAsp 434  
Db 1340 CGCCACAATGTCTCAATCCATCACTGCGACAGCGCATCGAGGAATCTCCGTCCAGGAC 1399  
Qy 435 LeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPhe 454  
Db 1400 CTGGCCTTGACCCCGCGGAAGCAGTATTGGCGCGAGTCTGGAGACGACCAACCGCTTC 1459  
Qy 455 ThrHisAla-----AspValThrAlaValProSerThrIlePheThrSerProVal 471  
Db 1460 ACCTTGGCAGCAGAAACGACCTCGGACATCGCGCAGACGCTCTGGGCAACGCCACA 1519  
Qy 472 PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleValArgAla 491  
Db 1520 TGGGCCACCTCGACGAGCGTCACTACGCGGGAACCTCGGTCAAGAGCGTCTCGCGCTC 1579  
Qy 492 GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 511  
Db 1580 GGCAACACCGCGCGCACCAACAG-----GTGGCCATCTCGTCCAGCGC 1624  
Qy 512 GlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAla 531  
Db 1625 GGCGCGAGTGGAGCATCGACTACGCGCGCGACAGCTCTGTGACCGCTCTCGCGGC 1684  
Qy 532 AlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValVal 551  
Db 1685 TATTGGCGGAGCGGACAGCATCTCTGTGACCGCTCTCGCGGC-----GTG 1735  
Qy 552 TyrAlaValGlyPheGlyAsnSerTrpAlaSerGlnGlyValProAlaAsnAlaGln 571  
Db 1736 CAGCGCTCGCAGTTCCAGGGCAGCTTTCCTCCGTCTCGAGCTGCGCGCGCGCGCTC 1795  
Qy 572 IleArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyr 591  
Db 1796 ATCGCTCGGACGAAGAACCAACAGCGTCTTCTACGCGGCTCTCGGATCGACCTTTTAC 1855  
Qy 592 ArgSerThrAspGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGly 611  
Db 1856 GTCAGCAGGACACCGGACGAGCTTC-----ACGCGGGG---CCCAAGCTGGG 1903  
Qy 612 AlaValGlyValMet-----PheHisAlaValProGlyLysGluGlyAspLeuTrpLeu 629  
Db 1904 AGCGCAGGACGATCCGGATATCGTCTCAACCGGACCCAGCGCGGACAGTGTGTATGTC 1963  
Qy 630 AlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIle--- 648  
Db 1964 TCGACGAGCTCGGATATTTTCCTTCACAGACTCGGCGACGACCTTGGCCCAAGTCTCC 2023  
Qy 649 ThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyr 668  
Db 2024 ACGCGCTGACCAACACCTACAGATCGCCCTGGGTGGGTCA---GGCTCGAAGTGG 2080  
Qy 669 ProAlaValValValGlyThrIleGlyValThrGlyAla-----TyrArgSer 686  
Db 2081 ---AACCTGTATGCTTCGGCACC-----GGCCCGTCAAGGCGTCCGCTCTACGCCAGT 2131  
Qy 687 AspAspCysGlyThrThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrp 706

Db 2132 GGAGACACCGCGCCTCTCTGGACGAGCATTCACAGGCTCCACGGGCTTCGGCTCCATCGAC 2191  
Qy 707 GlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgValTrpIleGlyThrAsnGly 726  
Db 2192 AGCACCAGGTGCGCGGAGCGGACGCGGACGCGGCGGCAAGTCTACGTGGGACCAACGGC 2251  
Qy 727 ArgGlyIleValTyr-----GlyAspIleGlyAlaProSerGly 740  
Db 2252 CGGGGCGTCTTTTACGCTCAGGGAACCGTCCGCGCGGCGCACGCGCGG 2299

## RESULT 7

US-10-156-761-2561  
; Sequence 2561, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 2561  
; LENGTH: 2217  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(2217)  
US-10-156-761-2561

Alignment Scores:  
Pred. No.: 3,09e-125 Length: 2217  
Score: 1405.00 Matches: 308  
Percent Similarity: 54.57% Conservative: 104  
Best Local Similarity: 40.79% Mismatches: 257  
Query Match: 34.81% Indels: 86  
DB: 15 Gaps: 21

US-09-917-376-3 (1-740) x US-10-156-761-2561 (1-2217)

Qy 6 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 25  
Db 124 TACCGTGGCGCAACGCCGCTCATCGGGGCGACCGGCTTCGTACCGCGGTCTTCCAC 183  
Qy 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAsp 45  
Db 184 CCTCGGTACGCGGCTTCGCTACGCCCGGACCGATCGCGCGCGCTACCGCTGGGAC 243  
Qy 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr 65  
Db 244 GACCGGGCGCGCTGGACCGCGCTCATCGACCACTCGGCTGGGACGAGTGAACCTC 303  
Qy 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85  
Db 304 CTCGGCGTCAAGCGATGCGCTCGACCCCGCACCGCGCGGCTCTACCTCGCGGTG 363  
Qy 86 GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGln 105  
Db 364 GGCACCTACCGCAGTGTGGGGGGGCAAC---GGCGGCTCTCGCTCCGAGGACCGC 420  
Qy 106 GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArg 125  
Db 421 GGGCCACCTGGAACCGGACCGACCTGACCGTGAAGCTCGCGCGGCAACGAGGCGCGC 480



126 GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 145  
1481 GCGCGCGGTGAGCGACTCTCTGTCAGCCGCGCGACGACACCTCTGGTGGGACG 540  
146 ProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 165  
541 ---CGGCACGACGGGCTCTCAAGTCGACCGACCGGGCGGCACCTTGGGACGCGGAC 597  
166 AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrGlyTyrGlnSer 185  
598 GCCTTCCCG-----GCGAAGCGCGAATCC 621  
186 AspileGlnGlyValValTyrValAlaPheAspLysSerSerSerLeuGlyGlnAla 205  
622 TCCGGGACGGAGTCTGTTCTC-----CTGTCGCCGCC 654  
206 SerLysThrIlePheValGlyValAlaAspProAsnAsnPro-----Val 220  
655 GGGCGCACCGTCTACCGCGCTGGGGTGACGCGGACGGCACCTCGGGCACGCGGAACCTG 714  
221 PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 240  
715 TACCACGCGCGAC---GGCACGACCTGGGGGCGCTCCCGCGCGCGCTCCGCGCAC 771  
241 -----IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAla 257  
772 TCCGCCAAGGTCCCGCTCCCGCGCGGTACGACGACGACACCGCGGAGCTGTACGTGACG 831  
258 ThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTyrLysPheSerVal 277  
832 TACGCGGACGACCG 891  
278 ThrSerGlyThrTrpThrArgIleSerProVal-----ProSerThrAspThr 293  
892 GCCACGGGACGTCGACCGGAGTACCCCGGTGAAGCGCGCGCGCGCGCGCGCGCGCGCG 951  
294 AlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIle 313  
952 TCGCGGACACCTTCCTGCTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1011  
314 MetValAlaThrGlnIleSerTrpProAspThrIleIlePheArgSerThrAspGly 333  
1012 GTCGTCTCCACCAACACCGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1071  
334 GlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnAspSerLeuArgTyr 353  
1072 GCGCGTACTGGACGCTCCCTCAAGGAC-----GCGCGC 1104  
354 ValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProVal 373  
1105 GTGTTTCGACGTGTCGAGACTCCCTTCCTCGACTCGGGCGACGAC----- 1149  
374 ProSerProLysLeuGlyTyrMetAspGluAlaMetAlaIleAspProPheAsnSerAsp 393  
1150 ---AAGCCGAAGTTCGGTGTGGATCCAGGGCTCGCGGTGCGACCCGTACGACTCCCG 1206  
394 ArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAsp 413  
1207 CACGTGCTGTACGGGACCGCGCGACCTCTACGCGACCCGCGGACCTCAAGCGCTGG--- 1263  
414 SerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsn 433  
1264 -----GCACCGCGGATCCGCGCGCTGGAGGAGCGCGCTGGCG 1302  
434 AspLeuIleSerProSerGly---AlaProLeuIleSerAlaLeuGlyAspLeuGly 452  
1303 CAACTGATCTCCCGCGGTGGGAGGACACCTGATCAGCGGACTCGGGACATCGGT 1362  
453 GlyPheThrHisAlaAspValAlaValProSerThr---IlePheThrSerProVal 471  
1363 GTGATGTACCAGCGGCTCACGGGCTCTCGTCGCGGGGATCGCGACGAAACCGCTG 1422

472 PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla 491  
1423 TTCCGGTTCGGGACGCGGACTCGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1482  
492 GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 511  
1483 GGTGGGCGGACACCGCAACGCG-----GCGTACTCCACGACGCGC 1524  
512 GlyLysAsnTrp-----PheGlnGlySerGluProGlyGlyValThrThrGlyGlyThr 529  
1525 GGGCGGACCTGGCGCGCTTCGAGCGCCGACCGCGACATCGCCAGGACCGACCGGACCG 1584  
530 ValAlaAlaSerAlaAspGlySerArgPheValTrpAla-----ProGlyAsp 545  
1585 ATCGCACCAAGTTCGCGCGCGGACACTGCTGTGTGCTCTCGTGCACGCGGACGCG--- 1641  
546 ProGlyGlnProValValTyrAla-----ValGlyPheGlyAsnSerTrpAla 561  
1642 -----ACGACGTACGCGCGCCACCGCTCGACGCAACGCGCGGAGCTGCTGCC 1689  
562 AlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThr 581  
1690 GAGGTCTCTCTCCCGAAGCGCGCACACCGCGTCCGCGACCGCGCGCGCTCGACGCGC 1749  
582 PheTyrAla-----LeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThr 599  
1750 TTCTACGCTACGACTTCGACAAATGGAAGCTATACCGCGACACTGACAGTGGCGCTGC 1809  
600 PheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAla 619  
1810 TTCGCGCGCGTGGCGGAGTGGCC---TCCGCGCGACGACGCGGTTCAAGTGTGTCGCG 1866  
620 ValProGlyLysGlyAspLeuTrpLeuAlaIleSer---SerGlyLeuTyrHisSer 638  
1867 GCGCGGACGACGAGCGCGGACCTGTGGCTCTCCGCAATGAAGCGCGGCTCTACCGGTC 1926  
639 ThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGly 658  
1927 ACCGACGCGGAGACCTTCGCCAGGATCGACAGTGTGGGCTCGTACACCTCGCGC 1986  
659 PheGlyLysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGly 678  
1987 TTGGCAAGCG 2046  
679 GlyValThrGlyAlaTyrArgSerAspCysGlyThrThrTrpValLeuIleAsnAsp 698  
2047 ACCATCACCGCGCTTACCGCTCCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2106  
699 AspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArg 718  
2107 GACGCGCACGAGTGGGGTGGATCGCGGAGCGCGTCTGCTGCGTACCGCGCATCCACG 2166  
719 ArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAsp 733  
2167 GCGGTCTACTCGCCACCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGAG 2211

## RESULT 8

US-10-156-761-1/c  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089



; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; PRIOR FILING DATE: 2001-08-02  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 1  
 ; LENGTH: 9025608  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces avermitilis  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (4187715)  
 ; OTHER INFORMATION: a, t, c, g, other or unknown  
 US-10-156-761-1

Alignment Scores: 1.22e-120 Length: 9025608  
 Pred. No.: 1405.00 Matches: 308  
 Score: 54.57% Conservative: 104  
 Percent Similarity: 40.79% Mismatches: 257  
 Best Local Similarity: 34.81% Indels: 86  
 Query Match: 15 Gaps: 21  
 DB:

US-09-917-376-3 (1-740) x US-10-156-761-1 (1-9025608)

QY 6 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 25  
 Db 3161882 TACCGTGGCGAAGCCGCTCATCGGGGACCGGCTTCGTACCGGCGTGTCTTCCAC 3161823  
 QY 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArgTrpAsp 45  
 Db 3161822 CCTCGGTACGGGTCCTCCCTACGCCCGGACCGATCGGGGCGCTACCGCTGGAC 3161763  
 QY 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr 65  
 Db 3161762 GACCGGGCGCGCTGGACCGGCTCATCGACCTCGGCTGGGACGACTGGAACCTC 3161703  
 QY 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnIysValTrpAlaAlaVal 85  
 Db 3161702 CTCGGGTGAGCGATGCGCTCGACCCGACCGACCGCGGCGGCTCTACCTCGCGCTG 3161643  
 QY 86 GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerAspAsn 105  
 Db 3161642 GGCACCTACGCCGAGTCGTGGGCGGCAAC---GGCGGCTCTCGGCTCCGAGGACCGC 3161586  
 QY 106 GlyAlaThrTrpGlnIleThrProLeuProPheLeysLeuGlyGlyAsnMetProGlyArg 125  
 Db 3161585 GCGCCACCTTGACCGGACCGGACCTGACCGTGAAGCTCGGCGGCAACGAGGACGCGCGC 3161526  
 QY 126 GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 145  
 Db 3161525 GCGCGCGGTGAGCGACTCTCGTCGACCGCGCGACGACGACACCCCTCTGGCTGGGCACG 3161466  
 QY 146 ProSerGlyIysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 165  
 Db 3161465 ---CGGCACGACGGGTCTCAAGTCGACGCGCGGCGGCGGCTTGGGACGCGCGGACC 3161409  
 QY 166 AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSer 185  
 Db 3161408 GCCTTCCCG-----CGGAGGCGAAGTCC 3161385  
 QY 186 AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeuGlyGlnAla 205  
 Db 3161384 TCCGGGACGGAGTCTGTTC-----CTCGTCGCGCGC 3161352  
 QY 206 SerLysThrIlePheValGlyValAlaAspProAsnAsnPro-----Val 220  
 Db 3161351 GGGCGCACCTGTACCGCGCTGGGTGACGCGGACGCGGACCTCGGGCACCGGCAACCTG 3161292  
 QY 221 PheTrpSerArgAspGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 240  
 Db 3161291 TACCGACGCGCCGAC---GGCAGACCTGGGGGCGGCTCCCGCGGCGGCTCTCGGAC 3161235  
 QY 241 -----IleProHisIysGlyValPheAspProValAsnHisValLeuTyrIleAla 257

Db 3161234 TCCGCCAAGTCCCGCTCGCGCGCGGTACGACACGACACCGCGCGAGCTGTACGTGACG 3161175  
 QY 258 ThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerVal 277  
 Db 3161174 TACGGCGACGACCCCGCGCGCGCGGTCCGACGCGACGCGTGCACAGCTGCGGTACC 3161115  
 QY 278 ThrSerGlyThrTrpThrArgIleSerProVal-----ProSerThrAspThr 293  
 Db 3161114 GCCACGGGACGCTGGACCGAGGTACCCCGGTGAAGCGCGGCGGACGACGACGCGC 3161055  
 QY 294 AlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIle 313  
 Db 3161054 TCGGGCGACACCTTCGCTACGGCGGGTCCGCGTCCGACGCGCGCGCGCGCACCTC 3160995  
 QY 314 MetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGly 333  
 Db 3160994 GTCGTCTCCACCAACACCGCTGGCGCGGACGCGGACGACGCTCTCCGCTCCAGGACGCGC 3160935  
 QY 334 GlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyr 353  
 Db 3160934 GCGCGTACCTGGACGTCCTCAAGAC-----GCCGCC 3160902  
 QY 354 ValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProVal 373  
 Db 3160901 GTGTTGACGCTGTCGAGACTCCCTTCTCGACTGGGCGGACGAC----- 3160857  
 QY 374 ProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspPropheAsnSerAsp 393  
 Db 3160856 ---AAGCCGAAGTTCGCTGCTGGATCCAGCGCTCGCGGTCCGACCTACGACCTCCAG 3160800  
 QY 394 ArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAsp 413  
 Db 3160799 CAGCTGTTACGAGGACCGCGGACCTCTACGCHCCGCGACCTCAACCGCTGG--- 3160743  
 QY 414 SerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsn 433  
 Db 3160742 -----GCACCGGATCCGCGGCTCGGAGGAGAGCGCGTGGC 3160704  
 QY 434 AspLeuIleSerProSerGly---AlaProLeuIleSerAlaLeuGlyAspLeuGly 452  
 Db 3160703 CAACTGATCTCGCCCGCGTCCGGGAGGACACCTGTATGACGCGACTCGGGGACATCGGT 3160644  
 QY 453 GlyPheThrHisAlaAspValThrAlaValProSerThr---IlePheThrSerProVal 471  
 Db 3160643 GTGATGTACACGAGCGGCTTCACGGCGTCTCCGTCGCGCGCATGGCGACGACCCCGTG 3160584  
 QY 472 PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla 491  
 Db 3160583 TTCGGTTCGCGACGGGACTCGCGGCGCGCGCGCGCGCGGTATGTCTCGGACG 3160524  
 QY 492 GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 511  
 Db 3160523 GGCTGGGGGACACGCGCAACGCG-----GGGTACTCCGACGACGCGC 3160482  
 QY 512 GlyLysAsnTrp-----PheGlnGlySerGluProGlyGlyValThrThrGlyGlyThr 529  
 Db 3160481 GGGCGGACCTGGGCGGCTTCGAGCGCCACCGCGCATCGCCAGGACGACGCGGACCG 3160422  
 QY 530 ValAlaAlaSerAlaAspGlySerArgPheValTrpAla-----ProGlyAsp 545  
 Db 3160421 ATCGCCACCAAGTTCGCGCGCGCGGACACTGCTGTGTCTTCGTGCTGACATGGGACGCGC--- 3160365  
 QY 546 ProGlyGlnProValValTyrAla-----ValGlyPheGlyAsnSerTrpAla 561  
 Db 3160364 -----ACGACGTACGCGCGCCCGCTCGAGGACAAACGCGCGGAGCTGTGTC 3160317  
 QY 562 AlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThr 581  
 Db 3160316 GAGGTCTCTCTCTTCCGGAAGGCGCGCACACCGGTTCGCGGACCGCGGCGGATCCGACGCGC 3160257  
 QY 582 PheTyrAla-----LeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThr 599



Db 3160256 TTCTACGCCTACGACTTCGACAAATGGAACGCTATATACGCCAGCACTGACAGATGGCGCTTCG 3160197  
Qy 600 PheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValAlaGlyValMetPheHisAla 619  
Db 3160196 TTACGGCCCGTGGCGGAGACTGCC---TCCGCGACACGAGATTCAAGCTGGTTCGG 3160140  
Qy 620 ValProGlyGlySerGlyAspLeuTrpLeuAlaAlaSer---SerGlyLeuTyrHisSer 638  
Db 3160139 GCGCGGGAGCAAGCGGCGACCTGGCTTCCGCGAAATGGAACGGGCTCTACCGGTCC 3160080  
Qy 639 ThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGly 658  
Db 3160079 ACCGACGCGGGGACACCTTCGCCAGGATGACAGCTCTCGGCTCGTACACCCCTCGGC 3160020  
Qy 659 PheGlySerAlaProGlySerSerTrpProAlaValPheValAlaGlyThrIleGly 678  
Db 3160019 TTCGGCAGGCGGCGGACGCGCGGACCTACCGCGGATCTACAGGTGGCTTCGACGGAG 3159960  
Qy 679 GlyValThrGlyAlaTyrArgSerAspCysGlyThrThrTrpValIleAsnAsp 698  
Db 3159959 ACCATCACCGCGCTACCGCTCCGACGCGCGCGGACATGGTTCGGATCAACGAC 3159900  
Qy 699 AspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArg 718  
Db 3159899 GACGCCACCAAGTGGGGTGGATCGCGGAGCGCGTGGTGCATCCCGCATCCACGCGC 3159840  
Qy 719 ArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAsp 733  
Db 3159839 CGGGTCTACCTCGCCACCACCAAGCGCGCGCATCCAGTACGGGAG 3159795

## RESULT 9

US-09-927-827-21  
; Sequence 21, Application US/09927827  
; Publication No. US20030036176A1  
; GENERAL INFORMATION:  
; APPLICANT: Bower, Stanley G.  
; APPLICANT: Ramseier, Thomas M.  
; TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris  
; FILE REFERENCE: 38-10(15824)B  
; CURRENT APPLICATION NUMBER: US/09/927,827  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: US 60/279,493  
; PRIOR FILING DATE: 2001-03-28  
; NUMBER OF SEQ ID NOS: 69  
; SEQ ID NO 21  
; LENGTH: 3668  
; TYPE: DNA  
; ORGANISM: Xanthomonas campestris  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1001)..(2668)  
US-09-927-827-21

Alignment Scores:  
Pred. No.: 2,866-99 Length: 3668  
Score: 1139.00 Matches: 267  
Percent Similarity: 50.85% Conservative: 120  
Best Local Similarity: 35.09% Mismatches: 295  
Query Match: 28.22% Indels: 79  
DB: 10 Gaps: 21

US-09-917-376-3 (1-740) x US-09-927-827-21 (1-3668)

Qy 1 AlaThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAsp 20  
Db 1217 GCCAGTCCGGGCCCTTACAGTGGCGCAGTGCAGTTCGCGCATTTGGCGTGGCGCTTTGTCCAC 1276  
Qy 21 GlyIleValPheAsnGlyAlaProGlyIleLeuTyrValArgThrAspIleGly 40  
Db 1277 GGTGTGCTGTTTCATCCCGCCGCAACGTGTCTGGGCTATGCGCGCACCGATGTGGTGGC 1336  
Qy 41 MetTyrArgTTrpAspAlaAlaAsnGlyArgTTrpIleProLeuLeuAspTrpValGlyTrp 60

Db 1337 CGGTACCCTGGGATGCCAGCGCAGCAGTGGACCGCGCTGACCGACTGGTTGGCGCT 1396  
Qy 61 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 80  
Db 1397 GACGACTGAACCTGATGGCATCGACGATTCGCGTGCAGCCCGCCGATGCCGATGCG 1456  
Qy 81 ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 100  
Db 1457 CTGTATCTGGCGCGCGCACCTATATGATGAA---CGCGCGGCGCACTCCGCGAGTGTG 1513  
Qy 101 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGly 120  
Db 1514 CGCTGTTCAACCGCGCGCGCATCTTCGACGCTGGCGACTGCGCTTTAAGCTGGGTGGT 1573  
Qy 121 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 140  
Db 1574 AACAGCTGGCGCGCGCCCAATGGCGAGCGGCTGGCGGTGGACCCGACGATGGCCCGCTG 1633  
Qy 141 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThr 160  
Db 1634 CTGCTGCTGGGCTCGCGCGATGCC---GGCTGTGGCGTAGCGACGATCGCGCGCGCAC 1690  
Qy 161 TrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 180  
Db 1691 TGGGCGAAGTGGCGCTTTCCGACGCGCGCTGGCGCGCTGCCACCGCGCGCAATCAT 1750  
Qy 181 ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSer 200  
Db 1751 GTTGGCGCGGACGACGCGGTG---GGATCGCTTGTGTCTGTTTCGACGCGCATGGC 1807  
Qy 201 SerLeuGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal 220  
Db 1808 AACACGCGCTCGCCCAACCGCGCATCTACGTGGCGTGTCCACCGACAGACCGCTG 1867  
Qy 221 PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 240  
Db 1868 TATGTGTCGAAGATCGCGCGCGAGTTGGGACCGCTGGCGGCAACCGCGCGCTG 1927  
Qy 241 IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn 260  
Db 1928 CGCCCGAGCCACATGGCGCGCGCAGCGATGGGCAC---TGGTATCTGAGCTATGGCGAC 1984  
Qy 261 ThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGly 280  
Db 1985 CAGCCCGCGCGGACCTGATGCGCGGGAGCGCTTGTGGAAATTCACCGCGCGCACAGGG 2044  
Qy 281 ThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr 300  
Db 2045 CGCTGGCGTGAATCAGCCCGATTCGCGAG---CCAGCCAGTGGCGCATGGATTCGGCTGG 2101  
Qy 301 SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer 320  
Db 2102 GGTGGGTGGCGGTGGATCCGCAACATCCGAGGTGCTGTGGCGCAGCACTTCGCGCGT 2161  
Qy 321 TrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIle 340  
Db 2162 CGCAGCGCGCGGACGAGCTGTATCGCAGCGTGGATGGTGGCAAGCACTGGACCGCTTG 2221  
Qy 341 TrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGlu 360  
Db 2222 -----TTGGCGGATGCGGTGTTCGATCAGACGCGCGCG 2254  
Qy 361 ProTrpLeuThrPheGlyValGlnProAsnProValProSerProLysLeuGlyTrp 380  
Db 2255 CGGTG---ACCGCATGCGCGCGCAC-----TGG 2284  
Qy 381 MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGly 400  
Db 2285 ATGGGG---GCCTGGGATCGATCCGCTTCGACGGCAACCATGCGCTGTTCGACCGCGC 2341  
Qy 401 AlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIle 420  
Db 2342 TACGGCATCTGGGCTCGCGCAATCTG-----CAGGATTTCCGC 2380



```
QY 421 AlaPro-----MetValLysGlyLeuGlucluThrAlaVal 432
DB 2381 GCACCGCGACGCCGCTCAGTGGTTCAGAGACCGTGGGCTGGAAGAACCGTGGCG 2440
QY 433 AsnAspLeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGly 452
DB 2441 CTGGACCTGCTCAGCCGATGCTGGCGGCATCTGCTCAGCGCGCTCGGCATATCGAC 2500
QY 453 GlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPhe 472
DB 2501 GGCTTCGCGCATGACGACTGACCGCGTG---CAGTTGCGAGTACGCGCGCCCGCGGTG 2557
QY 473 ThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGly 492
DB 2558 ACCAATGGCAAGCATGATCCCGCCGACGCGCGCGCGCGCGCGCGCGCGT 2617
QY 493 SerPhe-AspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyG1 512
DB 2618 ACCGTGGCGCGCGCGCGCAACACGAATCCGCGCGC---TGTATTACGCGTGTGGCGG 2674
QY 512 YIyAsnTrpPheGln--GlySerGluProGlyGlyValThrThrGlyGlyThrValAl 531
DB 2675 CAAGCAATGACTGCTGCTGGAGTGAGCGCGCGCGCGCGCGCGCGCGCGCATTCG 2734
QY 531 aLaSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValVa 551
DB 2735 CATTGTGCGATGCGCGCGCAGTGTGTGGCAGCGCGCGCGCGCGCGCAAT-----TG 2788
QY 551 lYrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaG1 571
DB 2789 GCGCAGTCCGACTTTCGCGCGCGCGAGTGGCGGTGGAACGCCCTGCGCAACCCCGGTG 2848
QY 571 nIleAsrSerAspArgValAsnProIyThrPheTyrAla-----LeuSerAsnGlyTh 589
DB 2849 GGTGTGACCGATCGGTGGAGCGCGCGCGCTGTTATGGTGGTGTGCGCGCGCGCGCA 2908
QY 589 rPheTyrArGserThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeu----- 607
DB 2909 GCTGTACGAGACACCGATGCGCGCGCGAGTTTCGCT-----GCGACCGGTGTACAGT 2962
QY 608 -----ProSerSerGlyAlaVa 613
DB 2963 GGGCAGCCCGCGCGCGATGAGCGCACCGCGCGCAGTGCCTCCGACCCGCTGGCGCGC 3022
QY 613 lGlyValMetPheHisAlaValProGlyLysGlyAspLeuTrpLeuAlaAlaSerSe 633
DB 3023 AGCGGTGGTGTATCTGCGCGAGCCCGCGCAAG-----ProSerSerGlyAlaVa 613
QY 633 rGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSe 653
DB 3054 -GGTGTGATGCGGTGGCAGGACGGT-----GGCTGCGAGTGTCTCTCAGCGCGCGCA 3106
QY 653 rAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheVa 673
DB 3107 AGCACGCTGCTGGCATCGGCAAGCGCTGGCGCGCGGTGGCGCGCGCGCGGTGTATCT 3166
QY 673 lValGlyThrIleGlyGlyValThrGlyAlaTyrArgSerAspCysGlyThrThrTr 693
DB 3167 GGGCGCGCGGTGACGGCGGTGATGGCTGTTCCTGCTCCGAGATGGCGCGCGCGAGTG 3226
QY 693 pValLeuIleAsnAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAs 713
DB 3227 GCAGCGCATCAGATGACCGCGCACCGCTTCGGCGC---CCGTACAGCGTGACCGGTGA 3283
QY 713 pHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAs 733
DB 3284 TCCGCGCATTCGCGCGGTGTGTACTTCGCCACCGCGCGCGCGCGCATTTTCTACGCGCA 3343
QY 733 p 733
DB 3344 T 3344
```

## RESULT 10

```
US-10-395-241-13
; Sequence 13, Application US/10395241
; Publication No. US20040038367A1
; GENERAL INFORMATION:
; APPLICANT: YAOI, Katsuro
; APPLICANT: MITSUISHI, Yasuishi
; TITLE OF INVENTION: NOVEL XYLOGLYCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE
; TITLE OF INVENTION: ENCODING THE SAME, AND METHOD OF PREPARING THE ENZYME
; FILE REFERENCE: Q73756
; CURRENT APPLICATION NUMBER: US/10/395,241
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: JP 2002-83433
; PRIOR FILING DATE: 2002-03-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 2367
; TYPE: DNA
; ORGANISM: Geotrichum sp. M128
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: (1)..(2367)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2367)
US-10-395-241-13
```

## Alignment Scores:

Pred. No.:	2,18e-97	Length:	2367
Score:	1117.00	Matches:	281
Percent Similarity:	50.37%	Conservative:	124
Best Local Similarity:	34.95%	Mismatches:	279
Query Match:	27.68%	Indels:	120
DB:	18	Gaps:	31

US-09-917-376-3 (1-740) x US-10-395-241-13 (1-2367)

```
QY 6 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 25
DB 10 TAGCAGTTCAAGAATGTCGCGATCGCGCGCGGGGTACATTACCGGGATTGTCGCGCAC 69
QY 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArgTrpAsp 45
DB 70 CCAAGACCAAGACCTGCTGTACGCGCGCGCAGCATTCGCGCGCGGTACCGTGGGAC 129
QY 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr 65
DB 130 GCAGGACGCTCCAAGTGGATCCCGCTCAACGACTTTATCGAGGCGCAGGACATGAACATT 189
QY 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85
DB 190 ATGGGCACCGAGTCGATCGCGCTCGAGCCCAACACCCGACAGCTGTACCTCGCGCAG 249
QY 86 GlyMetTyrThr---AsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAsp 104
DB 250 GGGCGTATGTCGGCGACGAGTGG-----CGCGGTTCTATGTCTCGAAGAC 297
QY 105 GlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGly 124
DB 298 CGCGGCGAGTCTGTACATCTACGAGTCGCGGTTCCTCGATGGGCGCAACGACATGGGA 357
QY 125 ArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGly 144
DB 358 CGCAACAATGGCGAGCGCTCGCTGTCAACCCGTTCACTCAACGAGGTGTGATGGGT 417
QY 145 AlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThrTrpSerGlnMet 164
DB 418 ACOCGTACA---GAGGGTATCTGGAAGATTCGAGACCGCGCGCAAGACTGGACAAACGTC 474
QY 165 ThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGln 184
DB 475 ACGTCCATCCGAGCGGTTTC-----ACCAACGGTATCGGATACACG 516
```







```

; LENGTH: 2646
; TYPE: DNA
; ORGANISM: Geotrichum sp. M128
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (120)..(2558)
US-10-395-241-11

Alignment Scores:
Pred. No.: 2,51e-97 Length: 2646
Score: 1117.00 Matches: 281
Percent Similarity: 50.37% Conservative: 124
Best Local Similarity: 34.95% Mismatches: 279
Query Match: 27.68% Indels: 120
DB: 18 Gaps: 31

US-09-917-376-3 (1-740) x US-10-395-241-11 (1-2646)
QY 6 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 25
DB 198 TAGGAGTTCAAGAAATGTCGATCGCGCGCGGGGTACATTACCGGGATTGTCGGCAC 257
QY 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArgTrpAsp 45
DB 258 CCAAGACCAAGACCTGCTGACGCGCGCACGGACATTGGCGCGGTACCGTGGGAC 317
QY 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTyrPAsnAsnTrpGlyTyr 65
DB 318 GCAGGCAGCTCCAAAGTGGATCCGCTCAACGACTTTATCGAGCGCGAGCATGAACATT 377
QY 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85
DB 378 ATGGGCACCGAGTCGATCGCTGGACCCCAACACCCCGACGAGGTGTACCTCGCGCAG 437
QY 86 GlyMetTyrThr---AsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAsp 104
DB 438 GGGCGGTATGTCGGCACAGTGG-----GCCGGGTCTATGTGTCGAGAC 485
QY 105 GlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGly 124
DB 486 CGCGGCCAGTCGTTTACATACGAGTTCGCGGTTCCCGATGGCGCCCAACGACATGGGA 545
QY 125 ArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGly 144
DB 546 CGCAACAATGGCGAGCGCTCGTGTCAACCCGTTCAACTCGAACCAGGTCTGGATGGGT 605
QY 145 AlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThrTrpSerGlnMet 164
DB 606 ACGCGTACAA---GAGGGTATCTGGAAGTTCGGACCGCGGCCGACCTGGACAAACGTC 662
QY 165 ThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGln 184
DB 663 ACCTCATCCGAGCGCGTTC-----ACCAACGGTATCGGATACACG 704
QY 185 SerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGln 204
DB 705 TCG-----GTCAATTTTCGACCCC-----GAA 725
QY 205 AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArg 224
DB 726 CGTAATGGGACCATCTACGCGAGCGGACTGCCCCCGAGGGC---ATGACGTACGCGAC 782
QY 225 AspGlyGlyAlaThrTrpGlnAlaValProGlyValaProThrGlyPheIle----- 241
DB 783 GACGGCGGTGTCGTTGGGAGCCAGTGGCGGGCCAGCGCTCCAGCTGGCTCAACAGGACC 842
QY 242 -----ProHisLysGlyVal 246
DB 843 ACGGGCGGTTCCCGGACAAAGACCGCGTCGATCGCGCGCGCAGCCCATGAAGTCGCT 902
QY 247 PheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAsp 266
DB 903 CTCACCCCC-----AACTTCCTTACGTGACTTACCGCGACTACCCCTGGTCCATGGGGC 956

```

```

QY 267 GlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArgIle--- 285
DB 957 GTCAGTTCGGCGGAAGTGTGGCGCAGAACCGCACCTCGGGCGCTGGGACGACATTACT 1016
QY 286 -----SerProValProSerThrAspThrAlaAsnAspTyrPheGly 299
DB 1017 CCCCCGCTCGGCAACTCGTCCGCTGCCCTCAACAACAGACAGCTTCCCTCGGGCGGA 1076
QY 300 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 319
DB 1077 TTTTGGGCTCAGCGTCAGCGCAGCGCAACCCCAACCGCTCTGCTGTCATCACC---CTC 1133
QY 320 SerTrpTrpProAspThrIle-----IlePheArgSerThrAspGlyGlyAlaThr 336
DB 1134 GACCGGACCCCGGACCCCGCTCGACAGCATCTACCTCTCAACCGATGCGGGCGGACC 1193
QY 337 TrpThrArgIleTrpAspTrpThrSer-----TyrPro 347
DB 1194 TGGAGGAGCTCACCCAGCTCTCGTCCCGTCCAACTCGAAGGTAACTGGGGCCACCCG 1253
QY 348 AsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProThrLeuThrPheGlyVal 367
DB 1254 ACTAAGCGCGCGGTAC---AAGGACGGCACCGCTTCCGTGGCTCGACTTCAACAAC 1310
QY 368 GlnPro-----AsnProValProSerPro-----LysLeuGly 379
DB 1311 GGTCCCACTGGGGGATACGGTGGCGCGCAGCTAGCCCGCGCTCACCAAGTTTGGC 1370
QY 380 TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr 399
DB 1371 TGGTGGATGAGCGTGTGCTTATCGATCCGTTCAACCCCGACGACCTGTGTAGCGCAG 1430
QY 400 GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLys-----TrpAspSerGly 415
DB 1431 GGGCGCACCATCTGGCGCGACCGACGCTCTCCGCTGTCGAGAGGACTGG----- 1481
QY 416 GlyGlnIleHisIleAlaPro-----MetValLysGlyLeuGluThrAla 431
DB 1482 -----CGCGCGAGCTGCTACCTCCAGATCCAGTATCGAGGAGATGCG 1526
QY 432 ValAsnAspLeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeu 451
DB 1527 ATCTCTGCTCGCTCGCTCGCCAAAGAGCGCGCGCTCTCTGTCGGGCATCGGTGACAT 1586
QY 452 GlyGlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProVal 471
DB 1587 AGCGGCATGAAGCACGACGACCTCACCAAG---CCCCAGAAAGATGTTTGGTGGCCCCCAG 1643
QY 472 PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla 491
DB 1644 TTCTCCAACTCGACAGCATCGACGCTGCGGCAACTTCCCAACGTTGCTCGCGGCC 1703
QY 492 GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 511
DB 1704 GGATCTCGGACACAGTACGACGCGCGTGGCGCGGTGGTGGTACGCGACTGACGCG 1763
QY 512 GlyLysAsnTrp-----PheGlnGlySerGluProGlyGlyValThr----- 526
DB 1764 GGAGACCGTGGACCATCTTCCCTCTGCTGCTGTAACGCGGACGACTACCGAC 1823
QY 527 GlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro-----Gly 544
DB 1824 GGCAGCATGTCAGTTCGACGCGGCGGCGAGCATGCTGTGTGTGTGTCGCCAAGCTTGAC 1883
QY 545 AspProGlyGlnProValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGln 564
DB 1884 GAGCAGGCTCGGGACCGTGTACTCCACGACTATGGCAAGAGCTGTCT- 1934
QY 565 GlyValProAla-----AsnAlaGlnIleArgSerAspArgValAsn 578
DB 1935 ---GTTCCCGCTGGCGACCTGAAGGCCAGACTGCCCAATGTCTCTCGGACAAAGTCCAG 1991

```



Qy 579 ProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyVal 598  
 Db 1992 GATGGCAGTTCTACGCTACCGATGGCGGAAGTTCTTCGCTCGACGCGGGAAG 2051  
 Qy 599 ThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHis 618  
 Db 2052 TCGTATGCGCGCAAGGCGCGGACTTGTCACT-----GGCACATCGCTCATGCTGCC 2105  
 Qy 619 AlaValProGlyLysGluGlyAspLeuTrpLeu---AlaAlaSerSerGlyLeuTyrHis 637  
 Db 2106 GTGAACCTCGGTGGCGCGGAGCTCTGGGTGCTGTTCGAGGGCGGTCTTTCAC 2165  
 Qy 638 SerThrAsnGlySerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsn 656  
 Db 2166 TCGACCGACTTGGCGCTCGTTCACGAGGTAGTACCGCAACGCGACCTCGTAGC 2225  
 Qy 657 ValGlyPheGlyLysSer-----AlaProGlySerSerTyrProAlaValPhe 672  
 Db 2226 GTCGGCGCGCCCAAGTCCAGTCGCGCAAGAGGCTAGCGCGCTCGCGGCTCTTC 2285  
 Qy 673 ValValGlyThr-----IleGlyGlyValThrGlyAlaTyrArgSerAspCysGly 690  
 Db 2286 ATCTGGGCGCACCGACAGCTGGAAGCGACATCGGCTGTACCGCTCGACGACACGCGC 2345  
 Qy 691 ThrThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIle 710  
 Db 2346 AGCACTGGACGCGCTCAATGACGAGGACACAACTACTCGGCG---CCACCATGATC 2402  
 Qy 711 ThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyIleVal 730  
 Db 2403 GAGGCGCGACCCCAAGTCTACGGCGCGGTGTATCTAGCGACGAACGCGCGGTATCGTG 2462  
 Qy 731 TyrGlyAspIle 734  
 Db 2463 TACGCGGACCTT 2474

RESULT 12

US-10-395-241-17  
 ; Sequence 17, Application US/10395241  
 ; Publication No. US20040038367A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: YAOI, Katsuro  
 ; APPLICANT: MITSUTSUSHI, Yasushi  
 ; TITLE OF INVENTION: NOVEL XYLOGLYCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE  
 ; TITLE OF INVENTION: ENCODING THE SAME, AND METHOD OF PREPARING THE ENZYME  
 ; FILE REFERENCE: Q73756  
 ; CURRENT APPLICATION NUMBER: US/10/395,241  
 ; PRIOR FILING DATE: 2003-03-25  
 ; PRIOR APPLICATION NUMBER: JP 2002-83433  
 ; PRIOR FILING DATE: 2002-03-25  
 ; NUMBER OF SEQ ID NOS: 19  
 ; SOFTWARE: Patent in version 3.2  
 ; SEQ ID NO 17  
 ; LENGTH: 2481  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Xyloglycan Oligosaccharide-Degrading Enzyme with Histidine Tag  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)..(2481)  
 US-10-395-241-17  
 Alignment Scores:  
 Pred. No.: 4,51e-97 Length: 2481  
 Score: 1114.00 Matches: 281  
 Percent Similarity: 50.25% Conservative: 123  
 Best Local Similarity: 34.95% Mismatches: 280  
 Query Match: 27.60% Indels: 120  
 DB: 18 Gaps: 31  
 US-09-917-376-3 (1-740) x US-10-395-241-17 (1-2481)

Qy 6 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 25  
 Db 13 TACGAGTTCAAGAATGTCGATCGCGCGCGGGTACATATACCGGATTCGCGCAC 72  
 Qy 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAsp 45  
 Db 73 CCAAAGACCAAGGACCTGCTGTACGCGCGCACGACATTGGCGCGCGTACCGTGGGAC 132  
 Qy 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr 65  
 Db 133 GCAGCGACGTCCAAGTGGATCCCGCTCAACGACTTATCGAGCGCGACGACATGACATT 192  
 Qy 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaVal 85  
 Db 193 ATGGGACCGAGTTCGATCGCGTGGACCCCAACACCCCGACAGGCTGTACCTCGCGCAG 252  
 Qy 86 GlyMetTyrThr---AsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAsp 104  
 Db 253 GGGCGCTATGTCGCGCGACGAGTGG-----GCGCGCTTCTATGTGTCCGAAGAC 300  
 Qy 105 GlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGly 124  
 Db 301 CGCGCGCGAGTCTTCAATCTACAGTCGCGCTTCCGATGGCGGCCACACATGGGA 360  
 Qy 125 ArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGly 144  
 Db 361 GCAACAATGCGAGCGCTCGCTGTCAACCGTTCAACTCGAACGAGGCTCTGGATGGGT 420  
 Qy 145 AlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMet 164  
 Db 421 ACGGCTACA---GAGGGTATCTGGAAGAGTTGGACCGCGCCCAAGACCTGGCAACACGTC 477  
 Qy 165 ThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGln 184  
 Db 478 ACGTCCATCCCGAGCGCTTC-----ACCACGGTATCGGATACGACG 519  
 Qy 185 SerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGln 204  
 Db 520 TCG-----GTCAATTTTCGACCCC-----GAA 540  
 Qy 205 AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArg 224  
 Db 541 CGTAATGCGACCATCTACGCGAGCGCGACTGCGCGCGAGGGC---ATGTACGTACGCGAC 597  
 Qy 225 AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIle----- 241  
 Db 598 GACGCGGTGTCTCGTGGAGCCAGTGGCGCGCCAGCGCTCGAGCTGGCTCAACAGGACC 657  
 Qy 242 -----ProHisLysGlyVal 246  
 Db 658 ACGGCGCGTTCGCGACAGAAGCCGCTCGATCGCGCGCGCCATGAAAGTCGCT 717  
 Qy 247 PheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAsp 266  
 Db 718 CTCACCCCC---AACTTCCTTACGTGACTTACGCGCGACTACCTCGTGCATGGGCG 771  
 Qy 267 GlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArgIle--- 285  
 Db 772 GTCACGTTCCGCAAGTCTGCGCGCAGAACCGCACTCGGGCGCTCGGGACGACATTA 831  
 Qy 286 -----SerProValProSerThrAspThrAlaAsnAspTyrPheGly 299  
 Db 832 CCCGCGTCGCGCAACTCGTCCCTGCCGTACACAAACACAGACGCTTCCCTGGGCGGGA 891  
 Qy 300 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 319  
 Db 892 TTTTTCGCTCTCAGCGTCGACGCGACCAACCCCAACCGTCTCGTGCATCACC---CTC 948  
 Qy 320 SerTrpTrpProAspThrIle-----IlePheArgSerThrAspGlyGlyAlaThr 336  
 Db 949 GACCGCGACCGCGGACCGCGCTCGACAGCATCTACCTCAACCGATCCGCGCGGAC 1008  
 Qy 337 TrpThrArgIleTrpAspTrpThrSer-----TyrPro 347



Db 1009 TGAAGGACGTCACCCAGCTCTCGTCCCGCTCCAACTCGAAGGTAACTGGGGCCACCCG 1068  
 QY 348 AsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProThrLeuThrPheGlyVal 367  
 Db 1069 ACTAACGCGCGCGGTAC---AAGACAGCGACGCGCTGTTCGGTGGCTCGACTTCAACAAC 1125  
 QY 368 GlnPro-----AsnProProValProSerPro-----LysLeuGly 379  
 Db 1126 GTTCCCAAGTGGGGGGGATAGGTGGCGGCACGGTAGCCCGGCTCACCAAGTTTGGC 1185  
 QY 380 TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr 399  
 Db 1186 TGTGGATGAGCGCTGTCTTATCGATCGTTCAACCCCGACACCTGATGTACGGCAG 1245  
 QY 400 GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLys-----TrpAspSerGly 415  
 Db 1246 GGGGCGACATCTGGGCGACCGACACGCTCTCCCGTTCGAGAAGGACTGG----- 1296  
 QY 416 GlyGlnIleHisIleAlaPro-----MetValLysGlyLeuGluThrAla 431  
 Db 1297 -----GCCCGGAGCTGTGTACCTCCAGATCGACGGTATCGAGGAATGCG 1341  
 QY 432 ValAsnAspLeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeu 451  
 Db 1342 ATCTGTCTCGCTCGCTCCCAAGAGCGCGCGCTCTCTCGGGCATCGTGACATT 1401  
 QY 452 GlyGlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProVal 471  
 Db 1402 AGCGCATGAGACGACGACCTACCAAG---CCCCAAGATGTTGGTGGCGCCAG 1458  
 QY 472 PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla 491  
 Db 1459 TTCTCCAACTCGACAGCATCGACCTCGGGCAACTTCCCAAGATGTTGGTGGCGGCC 1518  
 QY 492 GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 511  
 Db 1519 GATCTCTCGGACACGATGACAGCGCGTGGCGCGGTGGTGGTGGTGGTGGTGGTGG 1578  
 QY 512 GlyLysAsnTrp-----PheGlnGlySerGluProGlyGlyValThr----- 526  
 Db 1579 GAGACGCGTGGACCATCTTCCCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1638  
 QY 527 GlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValThrAlaPro-----Gly 544  
 Db 1639 GCGACGACGATTCAGTGCACGCGAGCGCGACGACGATGCTGTGTGGTGGTGGTGGTGG 1698  
 QY 545 AspProGlyGlnProValValThrAlaValGlyPheGlyAsnSerThrAlaAlaSerGln 564  
 Db 1699 GAGCAGGCTCGGGACCGGTGGTACTCGCACGACTATGGCAAGCGTGGTCT----- 1749  
 QY 565 GlyValProAla-----AsnAlaGlnIleArgSerAspArgValAsn 578  
 Db 1750 ----GTTCCGCTGGCGACTGAGGCCAGACTGCAATGTGCTCTCGACAAAGTTCAG 1806  
 QY 579 ProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyVal 598  
 Db 1807 GATGACAGCTTCTACGCTACCGATGGCGCAAGTCTTCTGCTCGACCGCGCGGGAAG 1866  
 QY 599 ThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHis 618  
 Db 1867 TCGTATGCGCCCAAGGGCGCGGACTTGTCACT-----GGCACATCGCTCAAGCTGGC 1920  
 QY 619 AlaValProGlyLysGluGlyAspLeuTrpLeu---AlaAlaSerSerGlyLeuTyrHis 637  
 Db 1921 GTGAACCTCTGGTGGCGCGCAAGTCTGGGTGGCTGTTCGCCGAGGGCGGTCTCTTCCAC 1980  
 QY 638 SerThrAsnGlyGlySerSerThrAlaIle---ThrGlyValSerSerAlaValAsn 656  
 Db 1981 TCGACGACTTGGCGCTCGTTTACAGAGGTAGGTACCGCAACCGACCGCTCGTGAGC 2040  
 QY 657 ValGlyPheGlyLysSer-----AlaProGlySerSerThrProAlaValPhe 672

Db 2041 GTTCGGCGCCCAAGTCCAAAGTCGGACGCGCAAGAGGTAGCGCGCTCCGGGTCTTC 2100  
 QY 673 ValValGlyThr-----IleGlyGlyValThrGlyValAlaTyrArgSerAspAspCysGly 690  
 Db 2101 ATCTGGGACCGCACAGAGCTTGAAGCGACATCGGCTGTACCGTCCGACGACACGCG 2160  
 QY 691 ThrThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIle 710  
 Db 2161 AGCACTGGACGCGGTCAATGACGAGGACCACTACTCTCGGGC---CCCACCATGATC 2217  
 QY 711 ThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyAspGlyIleVal 730  
 Db 2218 GAGGCGGACCCCAAGGTCTTACGGCGCGGTATCTAGGACAGAACGCGCGGTATCGT 2277  
 QY 731 TyrGlyAspIle 734  
 Db 2278 TAGCGCGACCTT 2289  
 RESULT 13  
 ; Sequence 1, Application US/10420191  
 ; Publication No. US20040067569A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: No. US20040067569A1ozymes Biotech, Inc.  
 ; APPLICANT: Rey, Michael W.  
 ; APPLICANT: Zaretsky, Elizabeth J.  
 ; APPLICANT: Haas, Jeffrey A.  
 ; TITLE OF INVENTION: Polypeptides Having Xyloglucanase Activity And Nucleic Acids  
 ; FILE REFERENCE: 10210.200-US  
 ; CURRENT APPLICATION NUMBER: US/10/420,191  
 ; PRIORITY FILING DATE: 2003-04-18  
 ; PRIOR APPLICATION NUMBER: US 60/373,987  
 ; NUMBER OF SEQ ID NOS: 6  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 1  
 ; TYPE: DNA  
 ; ORGANISM: Tricoderma reesei  
 ; US-10-420-191-1  
 Alignment Scores:  
 Pred. No.: 7,2e-74 Length: 5698  
 Score: 878.50 Matches: 334  
 Percent Similarity: 29.92% Conservative: 119  
 Best Local Similarity: 22.06% Mismatches: 270  
 Query Match: 21.77% Indels: 793  
 DB: 18 Gaps: 52  
 US-09-917-376-3 (1-740) x US-10-420-191-1 (1-5698)  
 QY 7 ThrTrpSerAsnValAlaIle---GlyGlyGlyPheValAspGlyIleValPheAsn 25  
 Db 242 TCATGAGACAGCTCAAGCTCGGCGCGCGGCTTCTCCCGGCGCATCATCTTCAA 301  
 QY 26 -----GluGlyAlaPro----- 29  
 Db 302 AGTACCTTCTTTCAGTTTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 361  
 QY 30 -----GlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrp--- 44  
 Db 362 CCCAAGACAAAGCGGTAGCATATGACGACAGATATTTGGCGGCTGTACCGCTCAA 421  
 QY 45 -----AspAla 46  
 Db 422 GGGTCTCTGTTTCCGCATCGTATACGTGCTGTCTATACCGCGCGCATCGCGGATTC 481  
 QY 47 AlaAsnGlyArgTrpIleProLeuLeuAspTrpVal-----GlyTrp----- 60  
 Db 482 GCGGACGACTCATGAGACCGCGCTCACGATGGGATTTGCTGATAATGCGCGGTGGCACAAG 541  
 QY 61 -----AsnAsnTrpGly-----Tyr 65



Db 542 CGGCTGCTGAGTACCTGGCGCAGTGCCTACCCCTAACGACTATTATACGGCGCAGCGTGTTC 601  
Qy 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85  
Db 602 TGGGGCATCAGCGCTGTTGGCTTGTATCCGAGGAGCATCAAAAGGTGTATGCCGCGAGTG 661  
Qy 85 ----- 85  
Db 662 ACCCGTAGCTGCAGACAAACGGAACTAGGCGTCTCTGTAGTTTCCACATACGGCGTAC 721  
Qy 86 GlyMetTyrThrAsnSerTrpAsp----- 93  
Db 722 GGCATGTATACGAACAGCTGGTC-TGTGATGTCTCTCAGATCTAGACCTATGATTTGGACG 780  
Qy 93 ----- 93  
Db 781 GCCGTACATATGCTTGTTCGACAGACACTACAGAGAGTCTAGATCTGGATCTAACCTGC 840  
Qy 94 -----ProAsnAspGlyAlaIleLeuArgSerSerAspGln 105  
Db 841 CTGACATTGGCCATATATAGGGATCCGAGTAATGAGCCATCATTCGCTCGTACAGCCG 900  
Qy 105 ----- 105  
Db 901 GACTGTAAACGGTATATATCCCTAGGCTCATCTCGTAGTAAGCGAGCAGTCTGGCG 960  
Qy 106 GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArg 125  
Db 961 GCGCAACGTTGCTCTTCAACCACTTGCCCTTCAAAAGTCGGGGGTAAACATGCCAGGACGC 1020  
Qy 125 ----- 125  
Db 1021 CCGCGTTGCACAGGAAGTGTGAAACGGGAAGTTTCAGCCCCCATTTGTACGGTCTCGG 1080  
Qy 126 GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 145  
Db 1081 GGACCCGGAGAGCGTCTGGCTGTGATCCGCCAACTCCACATCACTACTTTTGGTGT 1140  
Qy 146 Pro----- 146  
Db 1141 CCTCGGCTCTCGCAGACCGCAGCTAGGCGGTTGAGTTGTAGTAGTAAGAAACCCAGA 1200  
Qy 147 ---SerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMet--- 164  
Db 1201 CGCTCAGGAAACGGCCCTCTGGAAGTCTACGACGGCGCGTACCTTTTCCAAAGTCTCG 1260  
Qy 165 -----Thr 165  
Db 1261 GCGAGTCTTTCGGGAGACCTTCAGATGCCTGCGCGCGCACTGGAAAGGTTCCAGAGC 1320  
Qy 166 AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGln--- 184  
Db 1321 TCGTTACGGCAACTGGGAGCTACATCCAGACCGGAGTGATTCACACGGCTACAAACAGC 1380  
Qy 185 -----Ser 185  
Db 1381 AGCAAGTGCCTTGACCCCTGCATGTAGGGTCTGGGCTCACTAAGGTTGCCGATGTTGTCG 1440  
Qy 186 AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeuGlyGlnAla 205  
Db 1441 GACAAGCAAGACTCATGTGGTTACGTTTCGACTCAACACAGCAGCAGCCCGGGGAGCC 1500  
Qy 205 ----- 205  
Db 1501 CTGTTGCTTCCTGAGTACACCAATGCAAGCTGAGTTGGTGGTCTCGTGGCCCCCTCGG 1560  
Qy 206 SerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrp----- 222  
Db 1561 ACGTCTCGTATCTTGTGGCAGCGCTGATACATCACTACATCACTGCTTCACTATGTGAGCAGC 1620  
Qy 223 -----SerArg 224

Db 1621 TGCAGAGCATAGAAACCAACCGTGGCGACTATTGTAGTGACGAAGTCAGATACACTCGTGC 1680  
Qy 225 AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHis--- 243  
Db 1681 AATCGCGCTCCAGTGGAGTGCTGTACCGGGGAGCCAGGGAATATCTTTCTCTACAAG 1740  
Qy 244 -----LysGlyValPhe 247  
Db 1741 TTACGGCGAGGTGCACCTCAGCATGSCCCCGTCGGTCCCTTTATGAAGAGAGTGTTC 1800  
Qy 248 -----AspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyPro 264  
Db 1801 GCGAAACTGCAGCCAGCAGAGAAGCCTTGTATCTGACCTATTCCGATGGCAGGCGCG 1860  
Qy 264 ----- 264  
Db 1861 CGTTTTCAGTGGTGTCTCTTCGGAAACATAGACTGGATGAAGGTACCGTGTCCCGGC 1920  
Qy 265 TyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArg 284  
Db 1921 TATGATGGCACACTTGGCTCAGTGTGGAGTACGACATTGCGAGGGGAACTTGGAAAGAC 1980  
Qy 285 IleSerProValProSerThrAspThr----- 293  
Db 1981 ATACTACCGTGTGAACCGAGTCACACCTCCATCTGTAAAGTCCCTTTGAACCTTTCTG 2040  
Qy 294 -----AlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAsp--- 306  
Db 2041 ATCACCCTGCTCTGGATCAGATCTATCTTTGGCTTTGGCGGCTTGGCTCGATTTG 2100  
Qy 307 -----Arg 307  
Db 2101 TAGTGGGACAGAGACCTAGTCTAGATATGAACCCGGAACCCGCGAGCTAAAC 2160  
Qy 308 GlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAspThr----- 325  
Db 2161 CAAAAGCCAGGAACCCCTGTTGCTTCTTTGAACTCTTGTGGCGCAGATGCTCAGCTG 2220  
Qy 326 -----IleIle 327  
Db 2221 GTTTTCGGTCTTGGAAACAAACGAGAAGAACTTGAGAACCAACCGGTCTACAGTTCGAC 2280  
Qy 328 PheArgSerThrAspGlyAlaThrTrpThrArgIleTrpAspThrTrpSerTyrPro 347  
Db 2281 TTTGGTGCAGCGACTCTGGGACACATGGACCCGATCTGGCGTGGGCGACTATCCG 2340  
Qy 348 AsnArgSer---LeuArgTyrValLeuAspIleSerAlaGluPro----- 361  
Db 2341 AAAGCCAGCTGGCTGAGACC-CTGTTGTACCTCGGGCTAGACCCCGCCTCGATAGG 2399  
Qy 362 -----Tyr 362  
Db 2400 CACTGAGACCTATTACTACAGCATCTCAGTGAGTCACTTTACGATCCGATCCGATGG 2459  
Qy 363 LeuThrPhe-----GlyValGlnProAsn----- 370  
Db 2460 CTGACTCTGGATAATGATGTCTGATAGTCACTCAGTGAGATTTGTTAGGTACGCTTACC 2519  
Qy 371 -----ProValPro----- 374  
Db 2520 GACTGACTCTGCTTCAAGACTCCCAAAGCACCCGTGGATCAAGAAACAACATTTATCGATGTG 2579  
Qy 374 ----- 374  
Db 2580 ATGACTGAGACGAAGTTCTGAGGTTTCTGGCACCTAGTTCTTGTGTAATAAGCTACAC 2639  
Qy 375 -----SerPro-----LysLeuGlyTrpMetAspGluAlaMet 385  
Db 2640 TCGACGAGTCAACCGTCCGATGGTCTCATCAAGCGCCTCGGCTGGATGATTGAGTCTCTC 2699  
Qy 386 Ala----- 386  
Db 2700 GGCTCGCTCAGTGGCAGGCTACCAGAGTAGTTTCGGGAGCCGACCTACTAATCTCAGAGAG 2759



QY 387 ---IleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAla 405  
DB 2760 CAGATTGACCAACCGACAGCAACACCTGGGCTCTACGGCACCGGAATGACAATCTTTGGC 2819  
QY 406 Thr----- 406  
DB 2820 GTCTAACTGGGTGGCTGCTGCTGGTACCGAGATGCGGTGGCTTTACTGTTAGAAACG 2879  
QY 407 ---AsnAspLeuThrLysTrpAspSerGlyGlnIleHisIleAlaProMetValLys 425  
DB 2880 CGCCACGATCTCACCACATGGGACACGCGGCACCAATGTGTCAATCCAATCACTGGCGAC 2939  
QY 426 Gly----- 426  
DB 2940 GCGGTGCTAGTGGTGTGACCTGTGCGCGGTGTACAGTTAGTTAGTACCGCTCTG 2999  
QY 427 ---LeuGluGluThrAlaValAsnAspLeuIleSerProSerGlyAlaProLeuIle 445  
DB 3000 CGATCGAGGAATTCCTCGTCAGACCTGGCTCTGCACCCGCGGGAAGCGAGCTATTG 3059  
QY 446 Ser----- 446  
DB 3060 GCGTAGCTCTTAAGAGGCGAGTCTCTGGACCGGAGCGTGGCGGCTTCGCTCGATAAC 3119  
QY 447 ---AlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaVal----- 462  
DB 3120 CCGCAGTCGGAGACGACAAACGGCTTCACCTTTGGCAGCAGAAACGACCTCGGACATCG 3179  
QY 462 ----- 462  
DB 3180 GCGGTGACGCTCTGCTGTGCGGAGTGTACCGGTGGAGTGTCTTGTGAGGCGCTGTAGC 3239  
QY 463 ProSerThrIlePheThrSerProValPheThrGlyThrSerValAspTyrAla--- 481  
DB 3240 GCGCAGCGTCTGGGCAAGCCACATGGGCCACTCGACGAGCGTCTGACTAGCGCGG 3299  
QY 482 -----Glu 482  
DB 3300 AGCGTCTGCAGACCGTGTGCGGTGTACCGGTGGAGTGTCTCGCAGCTGATGCGGCGCC 3359  
QY 483 LeuAsnProSerIleIleValArgAlaGlySer----- 493  
DB 3360 TACTCGTCAAGAGCGTCTGCGCGTGGCAACACCGCGGACGCAACAGGTGGCCATC 3419  
QY 494 -----PheAspProSerGlnProAsnAspArgHisValAlaPhe----- 507  
DB 3420 TTGAGCCAGTTCTCGCAGCAGGCGGAGCGGCGGTGTGTGGCGCGCTGTTCACCGGTAG 3479  
QY 508 SerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThrThrGly 527  
DB 3480 ACGTCCGACGGCGCGCGAGTGGACATCGACTACGCGCGCGCACACGTCATGAACGGC 3539  
QY 528 Gly----- 528  
DB 3540 GGCAGGCTGCGCGCGCTGCACCTCGTAGCTGATGCGCGGCTGTGCAAGTACTTGGCG 3599  
QY 529 ---ThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGly 547  
DB 3600 CGCACGGTGGCTATTTCGGCCGACCGCGGACACGATCTCTGTGTGACCGGCTCGTCCGGC 3659  
QY 548 -----GlnPro 549  
DB 3660 GCGTGCCACCGGATAGCGGCTGCGGCTGTGTAGGAGACGAGTGTGGCGGACGAGCGCG 3719  
QY 550 ValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsn 569  
DB 3720 CTCGAGGCTCG---CAGTTCCAGGGCAGGTTTGGCTCTCGAGCCTGCGCGCGCGG 3776  
QY 570 AlaGln----- 571  
DB 3777 GCCGAGTCGCGAGCGCTCAAGGTTCCCGTGTGAACGAGGAGGAGCTCGGACGGCGCGCCG 3836

QY 572 ---IleArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThr 589  
DB 3837 CGGCTCATCGCTCGGCAAGAAGACCAACAGCGTCTTCTACGCGGCTCCGGATCGACC 3896  
QY 590 Phe----- 590  
DB 3897 TTTTAGTAGCGGAGCCTGTTCTTCTGGTTGTCCGAGAAGATGCGGCGAGCTAGCTGG 3956  
QY 591 ---TyrArgSerThrAspGlyValThrPheGlnProValAlaAlaGlyLeuProSer 609  
DB 3957 AAAACGTCAGCAAGACACCGGACAGCTTC-----ACGCGCGG---CCCAAG 4004  
QY 610 SerGlyAlaValGlyValMet----- 616  
DB 4005 CTGGGCGAGCGGAGTGCAGTCTGTCGCGCTGTGCGAAGTGTGCGGCGCGCGGTTG 4064  
QY 617 -----PheHisAlaValProGlyLysGluGlyAspLeu 627  
DB 4065 GACCCGTCGCTCCCTCGATCCGGGATATCGTGTCTCACCCGACCAACCGCGGACGTTG 4124  
QY 628 TrpLeuAlaAlaSerSer----- 633  
DB 4125 TATGTCCTGACCGACCGGTAGGCCCTATAGCGAGGTGGTGTGGCGCGCGCTGCAC 4184  
QY 634 -----GlyLeuTyrHisSerThrAsnGlyGlySerSer----- 644  
DB 4185 ATACAGAGTGGTGTCTCGCATATTCGCTCCACAGACTCGGCGACGACCTTTGGCCAA 4244  
QY 644 ----- 644  
DB 4245 GTCTCCACCGCCTGAAGCGGTATAGGCGGAGGTGTCTGAGCCGCTGCTGGAACCGGTT 4304  
QY 645 -----TrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSer 662  
DB 4305 CAGAGTGGCGGACTCCAACTACCATCGCTCGGCTGGTGGGCTCAGGCTCGAAC 4364  
QY 663 -----AlaProGlySerSerTyr 668  
DB 4365 TGGAACTGTATCGCTGGTTGTGGTGTCTAGCGGACCCACACCGAGTCCGAGCTTG 4424  
QY 669 ProAlaValPheValValGlyThrIleGlyValThrGlyAla-----TyrArgSer 686  
DB 4425 ACCTTGGACATACGGATCGGCACC-----GGCCCGTCAGGGGCTCGCTCTACGCCAGT 4478  
QY 687 AspAspCysGlyThrThrTrpVal----- 694  
DB 4479 GGAGACAGCGCGGCTCTGGAAGCGGTGGCGGCGAGTCCCGGCGGAGATCGGTC 4538  
QY 695 -----LeuIleAsnAspAspGlnHisGlnTyrGly----- 704  
DB 4539 CCTCTGTGCGCGGAGGACCTCGACATCCAGGCTCCAGGCTTCGGCTCCATCGAC 4598  
QY 704 ----- 704  
DB 4599 AGCACCAAGTCCCGGACGCGGCTGTAGTCTCCGAGGGTCCGAAAGCGAGGTAGCTG 4658  
QY 705 AsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArgValTyrIleGlyThr 724  
DB 4659 TCGTGGTTCCAGCGGCGGCTCGCGCAGCACCGCGCGG-----CAAGTCTACGTGGGACC 4712  
QY 725 AsnGlyArgGlyIleValTyrGlyAspIleGlyAlaPro 738  
DB 4713 AACGCGCGGCGGCTCTTTTACGCTCAGGCGTGTGGCGGCGCC 4754

## RESULT 14

US-10-653-047-7511  
; Sequence 7511, Application US/10653047  
; Publication No. US2004029367A1  
; GENERAL INFORMATION:  
; APPLICANT: Randy M. Berka  
; APPLICANT: Michael W. Rey  
; APPLICANT: Jeffrey R. Shuster  
; APPLICANT: Sakari Kauppinen



APPLICANT: Ib Groth Clausen  
APPLICANT: Peter Bjarke Olsen  
TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
TITLE OF INVENTION: Expression  
FILE REFERENCE: 5849.200-US  
CURRENT APPLICATION NUMBER: US/10/653,047  
CURRENT FILING DATE: 2003-08-29  
PRIOR APPLICATION NUMBER: US/09/533,559  
PRIOR FILING DATE: 2000-03-22  
PRIOR APPLICATION NUMBER: 09/273,623  
PRIOR FILING DATE: 1999-03-22  
NUMBER OF SEQ ID NOS: 7860  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 7511  
LENGTH: 1103  
TYPE: DNA  
ORGANISM: Tricoderma reesei  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)...(1103)  
OTHER INFORMATION: n = A,T,C or G  
US-10-653-047-7511

Alignment Scores:  
Pred. No.: 7.64e-28 Length: 1103  
Score: 392.50 Matches: 114  
Percent Similarity: 45.05% Conservatives: 36  
Best Local Similarity: 34.23% Mismatches: 125  
Query Match: 9.72% Indels: 58  
DB: 20 Gaps: 10

US-09-917-376-3 (1-740) x US-10-653-047-7511 (1-1103)

Qy 302 GlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrp 321  
Db 3 GGCCTGCGCTCGATTGCAAGCCAGGACCTTGTGTGCTTCTTGAACCTCTGG 62  
Qy 322 TrpProAspThrIleIlePheArgSerThrAspGlyAlaThrTrpThrArgIleTrp 341  
Db 63 TGGCCAGATGCTCAGCTGTTTCGTGCGACCGACTCTGGGCAACATGAGCGCCGCTGG 122  
Qy 342 AspTrpThrSerTrpProAsnArgSerLeuArgTrpValLeuAspIleSerAlaGluPro 361  
Db 123 GCGTGGGAGCTATCCAGCTGAGACCTTATTACTACAGCATCTCACTCCCAAGCACCG 182  
Qy 362 TrpLeuThrPheGly---ValGlnProAsnProProValProSer-----Pro 376  
Db 183 TGGATCAAGAACAACTTTATCGATGTGACGAGCGAGTCACCGTCGATGCTTNATCAAG 242  
Qy 377 LysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeu 396  
Db 243 CGCCTCGGCTGGATGATTGAGTCTNTCGATGATTGACCCACCGCAGCAANNACTGGCTT 302  
Qy 397 -TyrGlyThrGlyAlaThrLeuTrpAlaThrAsnAspLeuThrLysTrpAspSer---- 414  
Db 303 TTACGGACCCGGAGTACAAATNTTGGCGGCCACGANTTCCAACTGGACACCGGCC 362  
Qy 415 -----GlyGlyGlnIleHisIleAlaProMetValLys-GlyLeuGluThrAlaV 432  
Db 363 ACAATGTGTCAATCAATTAATTACTTGGCAGCGGATTCGAAGGAATT-----TTTCGT 416  
Qy 432 alAsnAspLeuIleSerProSerGly----- 441  
Db 417 TCAAGGACCTGGCCTTTTTCACCGGGGGAAGCGAGCTTTTGGCGCGAAGTCCGAGACG 476  
Qy 442 -----AlaProIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspV 459  
Db 477 ANCAACGGGTTTACTTTTGGCCAGCAGAAACGACCTTGGG----- 517  
Qy 459 alThrAlaValProSerThrPheThrSerProValPheThrThrGlyThr-SerVal 478  
Db 518 -----ACATTCGCGCAGACGGTTTGGGCAACTCCACATGGGCCACTCGACGAGCGTC 572

Qy 479 AspTyrAla--GluLeuAsnProSerIleIleValArgAlaGlySer----- 493  
Db 573 GACTACGCGCGGAACTCGGTCAAGAGCGTTCGTCGCGTTCGCAACACCGCGCGGACG 632  
Qy 494 -----PheAspProSerSerGlnProAsnAspArgHisValAlaPheS 508  
Db 633 CAACAAGGTGGCCATTTTTCGGAACGCGCGCGCGCGACGTNGGAA-----GCAATTC 686  
Qy 508 erThrAspGlyGlyLysAsnTrpPheGlnGlySerGlnProGlyGlyValThrThrGlyG 528  
Db 687 GAACCTAACGCT-----GGMTCCGAACACACCGTTTCCATTGGAACGCGC 728  
Qy 528 lyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyG 548  
Db 729 GCGCGTGGCTATTTCGCGCGCGCGCGACACGATCTCTGTCGACCGCTCGTCGCGC- 787  
Qy 548 lnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProA 568  
Db 788 -----GTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCTCCTCGAGCTTCGCGC 839  
Qy 568 laAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnG 588  
Db 840 CGGCGCGCTCATCGTCTCGGACAAGACCAACAGCGTNTTCTACGCGCGCTCGGAT 899  
Qy 588 lyThrPheTyrArgSerThrAspGlyGly 597  
Db 900 CGACCTTTTACGTACGCAAGGACACCGCGC 928

## RESULT 15

US-10-246-330-3  
; Sequence 3, Application US/10246330  
; Publication No. US20030166030A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Toole, George A.  
; APPLICANT: Mah, Thien-Fah  
; TITLE OF INVENTION: METHODS TO STUDY AND MECHANISMS OF  
; TITLE OF INVENTION: BIOFILM-SPECIFIC ANTIBIOTIC RESISTANCE  
; FILE REFERENCE: 14537-002001  
; CURRENT APPLICATION NUMBER: US/10/246,330  
; CURRENT FILING DATE: 2002-09-18  
; PRIOR APPLICATION NUMBER: US 60/323,241  
; PRIOR FILING DATE: 2001-09-18  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 7407  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(7404)  
US-10-246-330-3

Alignment Scores:  
Pred. No.: 7.2e-12 Length: 7407  
Score: 238.00 Matches: 205  
Percent Similarity: 32.59% Conservatives: 88  
Best Local Similarity: 22.80% Mismatches: 314  
Query Match: 5.90% Indels: 292  
DB: 16 Gaps: 50

US-09-917-376-3 (1-740) x US-10-246-330-3 (1-7407)

Qy 11 ValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGly 30  
Db 3250 GTCAACCTGAGCAATGGCAGCAGCTCAGCGC-----ACCGCGGAGCGCGC 3297  
Qy 31 IleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArg 50  
Db 3298 AGCAGCGGTGATTCTCACCAC-----GGCAACGGCAAT 3330  
Qy 51 TrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr----- 65



3331 CCGATCGCCGAGGTACCCGCGACGGCAGCGCAACTGGACCTACACCCCGTCCACGCCG 3390  
Qy 66 -----Aandlyl---valvaliserilealalaspProIleAenThrAsnlyValTrp 82  
Db 3391 ATCCCAACGGTACTGTGTCAACGTGTGTGCCAGGACGCTCCGTAACAGCAGCCCG 3450  
Qy 83 AlaAlaValGlyMetTyThrAsnSer -----TrpAspProAsnAsp 96  
Db 3451 CCGGCGCGGTGACCGTTCGATCCAGCGCGCGCGCGCGCGGTGATCAACCCGAGCAAC 3510  
Qy 97 GlyAlaIleLeuArg---SerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuPro 115  
Db 3511 GCGGTGTCATCAGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3561  
Qy 116 PheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspPro 135  
Db 3562 ---GATCGCGCGCGCAACCCGATAGCGGAGGTACACCGCGCGC- 3600  
Qy 136 AsnAsnAspAsnIleLeuTyThrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThr 155  
Db 3601 -----GGCAGCGGCACTGGAGCTTCACG 3624  
Qy 156 AspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrlleAla 175  
Db 3625 CCG -----GGCACCCCGCTGGCC 3642  
Qy 176 Asn-----ProThrAspThrThrGlyTyrlGlnSer 185  
Db 3643 AACGGCACGGTGTATCGTCCACGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3702  
Qy 186 -----AspIleGlnGlyValTrpValAlaPheAspLysSerSerSer 201  
Db 3703 GCCACCACGGTGTATCGTCCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3762  
Qy 202 Leu-----GlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsn---Asn 218  
Db 3763 ACCATCAGCGGACCGCGGAGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3822  
Qy 219 ProVal---PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaPro 237  
Db 3823 CCGATCGGGAACACACCG 3882  
Qy 238 -----Thrgly--- 239  
Db 3883 CTGGCCAAACGGCACGGTGTGTCAACGCGGTGGCGCCAGCACCCCTCGCGGCAATACCGCGCG 3942  
Qy 240 -----PheIleProHisLysGlyValPheAspProValAsn 251  
Db 3943 CAGGCGCAGCACTACCGTGGACCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4002  
Qy 252 HisValLeuTyrlleAlaThrSerAsnThrGlyGlyProTyrl-----AspGlySer 268  
Db 4003 GGCACCTGTCTCAACGGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4062  
Qy 269 SerGlyAspValTrpLysPheSerValThr---SerGlyThrTrp-----ThrArgIle 285  
Db 4063 GGCACCG 4122  
Qy 286 SerProValProSer-----ThrAspThrAlaAsnAspTyrlPheGlyTyrlSer 301  
Db 4123 TCGCAACTGCCCAACGGCACCGGTGTCAACGGTGGACCGCGCGCGCGCGCGCGCGCGCGCG 4182  
Qy 302 GlyLeu-----ThrIleAspArgGlnHisProAsnThrIleMetVal----- 315  
Db 4183 AGCCTTCGCGTACACGACGCGGTGATTCCTCGCTGCGGTTCGATCCCGCGAGTGGATCCG 4242  
Qy 316 -----AlaThrGlnIleSerTrpTrpProAsp-----ThrIleIlePheArgSer 330  
Db 4243 AGCAACGGTTCGTGTGTATCAGCGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4296  
Qy 331 ThrAspGlyGly-----AlaThrTrpThrArgIleTrpAsp 342  
Db 4297 ACCGATGGCAACGGCAACCCGATTTGGCCAGGTTCACCGCGCGCGCGCGCGCGCGCGCGCG 4356

343 TrpThr-----SerTyrlProAsnArgSerLeuArgTyrlVal 354  
Db 4357 TTCCTCCAGGCATCCGCTCGCGATGGCACCGTGTCAACGTGTGTGGCGCGCAGGCCA 4416  
Qy 355 LeuAspIleSerAlaGluProTrpLeu-----ThrPheGlyValGlnProAsnProPro 372  
Db 4417 AGCAATGTGCACAGTGCAGCGCGCGGTGATCACTGTGTGTGTGGCGCGCGCGCGCGCG 4476  
Qy 373 ValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSer 392  
Db 4477 GTG-----ATCGATCCGACGACCGC 4497  
Qy 393 AspArgMetLeuTyrlGlyThr-----GlyAlaThrLeuTyrlAlaThrAsnAspLeu 409  
Db 4498 ACCGAG---ATAAGCGGTACCGCGGAGCGCGCGCGCGCGCGGTGATCTCTCACC----- 4545  
Qy 410 ThrLysTrpAspSerGlyGlnIleHisIleAlaProMetValLysGlyLeuGluGlu 429  
Db 4546 -----GATGGCGCGGCAAC-----CCGATCGCGCGCAGCCACCGCGCGCGC 4584  
Qy 430 ThrAlaValAsnAspLeuIleSerPro-----ProSerGlyAlaProLeuIle 445  
Db 4585 GGCAGCGGCAACTGGACGTTCAACCGGAGCACCCCGTGGCCACCGCACCGTGTCAAC 4644  
Qy 446 SerAlaLeuGlyAspLeuGlyGlyPheThrHis-----AlaAspVal 459  
Db 4645 GCGTGGCGCCAGACCCCGCGCGCAATACAGCGGTCCGCGCACCGTCACTCGGTGATGCC 4704  
Qy 460 ThrAlaValProSerThrIlePhe-----ThrSerProValPheThrThrGlyThrSer 477  
Db 4705 ATCGCCCGCGCGCGCGGTGATCAATCCGAGCAATGGAGTCTGTATCATCAGCGGTACG-- 4761  
Qy 478 ValAspTyrlAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSer 497  
Db 4762 -----GCGAAGCGCGCGCGCGGTGATCTCACCGACGCGCAACCGCACCCGATC 4812  
Qy 498 SerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPhe--- 516  
Db 4813 GCGCAG-----GTCCCGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4851  
Qy 516 ----- 516  
Db 4852 AGCCCGCGCACCGCGTGGCCAAACGGCTCGGTGATCAATCGGTGGCGCGCGCGCGCGCGCG 4911  
Qy 517 ---GlnGlySerGluProGlyGlyValThr----- 525  
Db 4912 GGCACACACAGCAGTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4971  
Qy 526 -----ThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrp 541  
Db 4972 ATCGATCCGAGCAACCGTAGCTAGCTCGCGGTACCGCGCGCGCGCGCGCGCGCGCGCG 5031  
Qy 542 AlaProGlyAspProGlyGlnProVal-----ValTyrlAlaValGlyPheGlyAsnSer 559  
Db 5032 CTCACCGCGCGCAACCGCAACCGCATCGCGCAGGTACCGCGCGCGCGCGCGCGCGCGCG 5088  
Qy 560 TrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnPro 579  
Db 5089 TGGAGTTCACGCGCGCGCGCGCGCG----- 5112  
Qy 580 LysThrPheTyrlAlaLeuSerAsnGlyThrPheTyrlArg----- 592  
Db 5113 -----CTGTCCAATGGCACCGTGTCAATCGGTGGCGCGCGCGCGCGCGCGCGCGCG 5157  
Qy 593 ---SerThrAspGlyGlyVal-----ThrPheGlnProValAlaAlaGlyLeuPro--- 608  
Db 5158 GGCACACACG 5217  
Qy 609 -----SerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysGlyAsp 626  
Db 5218 ATCGACCGGAGCAACCGTGTTCGAACCTCAGCGCGCACCGCGCGCGCGCGCGCGCGCGCG 5277



Qy	627	LeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSer---ThrAsnGlyGlySerSerTrp	645
Db	5278	CTCACCAGTGGCAATCCGATCGGCCAGACCTTGGCCGACGGCAGCGCACTGG	5337
Qy	646	SerAlaIleThrGlyVal-----SerSerAlaValAsnValGlyPheGlyLys	661
Db	5338	AGCTTCACGCCGGGCACCGCGTGGCCAAACGGCACGGTGGTCAATGCCGTGGCCCGCAGGAC	5397
Qy	662	SerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGlyGlyValThr	681
Db	5398	CCGGCCGGCAATACACAGCGGCCGCCAGCACCGGTGGACACGGTGGCTCCGGCCACG	5457
Qy	682	GlyAlaTyrArgSerAspCysGlyThrThrTrpValLeuIleAsnAspAspGlnHis	701
Db	5458	-----CCGGTGATCAATCCAGCAAC----	5478
Qy	702	GlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgValTyr	721
Db	5479	-----GGCAGCGTGATCACCGCACCC---GCCGAGGTCCGGCCCAAGGTG	5520
Qy	722	IleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyGlyAlaProSerGly	740
Db	5521	ATCCTCACCGATGGCAACCGCAACCCGATCGCGCAGAGCACCGCGCGCAGCGAGTGGT	5577

Search completed: October 6, 2005, 04:45:37  
 Job time : 11542.5 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 07:48:48 ; Search time 17,0496 Seconds  
(without alignment)  
2018.911 Million cell updates/sec

Title: US-09-917-376-4  
Perfect score: 470  
Sequence: 1 VSGGVKVVQKNDSPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_16Dec04:.\*  
1: Geneseqp1980s:.\*  
2: Geneseqp1990s:.\*  
3: Geneseqp2000s:.\*  
4: Geneseqp2001s:.\*  
5: Geneseqp2002s:.\*  
6: Geneseqp2003as:.\*  
7: Geneseqp2003bs:.\*  
8: Geneseqp2004s:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	469	99.8	88	7	ADD22925 Acidother
2	469	99.8	89	7	ADD22924 Acidother
3	469	99.8	154	6	ABP73020 Amino aci
4	469	99.8	154	7	ADJ38294 A. cellul
5	469	99.8	762	6	ABP73022 Amino aci
6	469	99.8	762	7	ADJ38291 A. cellul
7	466	99.1	150	6	ADH36640 A. cellul
8	466	99.1	150	8	ADH36640 A. cellul
9	466	99.1	1228	6	ABP73029 Amino aci
10	466	99.1	1228	8	ADH36636 Acidother
11	462	98.3	88	6	ABP73018 Amino aci
12	462	98.3	88	8	ADJ38291 A. cellul
13	462	98.3	89	6	ADH36640 A. cellul
14	462	98.3	89	8	ADH36640 A. cellul
15	462	98.3	957	7	ABP73015 Amino aci
16	462	98.3	957	7	ADJ38291 A. cellul
17	462	98.3	957	8	ADJ38291 A. cellul
18	436	92.8	153	6	ABP71658 A. cellul
19	436	92.8	153	7	ADP75898 Carbohyd
20	436	92.8	1121	6	ADP71656 A. cellul
21	436	92.8	1121	7	ADP75895 Acidother
22	231.5	49.3	782	2	AAR15625 Cellulase
23	195.5	41.6	616	2	AAY13494 Truncated
24	195.5	41.6	616	5	Aae16325 Active ce
25	194.5	41.4	1751	2	AAY13493 Truncated

ALIGNMENTS

RESULT 1

ADD22925  
ID ADD22925 standard; protein; 88 AA.

XX  
AC ADD22925;

XX  
AC ADD22925;

DT 15-JAN-2004 (first entry)

XX  
DE Acidothermus cellulolyticus avicelase AVIII CBD III #2.

XX  
KW enzyme; Aviii; cellulose reduction; agricultural biomass;

KW municipal solid waste; glycoside hydrolase; avicelase.

XX  
OS Acidothermus cellulolyticus.

XX  
PN US2003108988-A1.

XX  
PD 12-JUN-2003.

XX  
PF 18-OCT-2002; 2002US-00155400.

XX  
PR 28-JUL-2001; 2001US-00917376.

XX  
PA (DING/) DING S.

XX  
PA (ADNE/) ADNEY W S.

XX  
PA (VINZ/) VINZANT T B.

XX  
PA (HIMM/) HIMMEL M E.

XX  
PI Ding S, Adney WS, Vinzant TB, Himmel ME;

XX  
XX WPI; 2003-810853/76.

XX  
DR New isolated thermal tolerant avicelase polynucleotide useful for

XX  
PT detection of a polynucleotide encoding Aviii and for reducing cellulose

XX  
PT in a starting material, e.g. municipal solid waste.

XX  
PS Claim 16; SEQ ID NO 5; 29pp; English.

XX  
CC The invention relates to an isolated polynucleotide molecule encoding a

XX  
CC thermostable Aviii polypeptide. The polynucleotide is useful for

XX  
CC detection of a polynucleotide encoding Aviii. The polynucleotide is

XX  
CC useful for reducing cellulose in a starting material which involves

XX  
CC administering to the starting material, e.g. agricultural biomass or

Aae16324 Active ce  
Aay13492 Truncated  
Aae16323 Active ce  
Aao55907 Bacillus  
Ads44721 Bacterial  
Aao22444 Protein e  
Aao22443 Protein e  
Aau98063 Bacillus  
Aar42122 NK-1 cell  
Aar13227 Novel end  
Aaw18790 Corrected  
Aar95080 Cellulose  
Adg14244 Cellulose  
Adg14258 Cellulose  
Aay54123 A manna  
Aay28850 Pectate 1  
Aay43218 Pectate 1  
Aaw5238 Scaffold  
Aaw43108 C. thermo  
Adj35112 Xylanase



Matches	88;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	1	VSGGVKQVQKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLVNCD	60						
Db	1	VSGGVKQVQKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLVNCD	60						
QY	61	WAAMCGCNIRASFGSVNPATPTADTYLQ	88						
Db	61	WAAMCGCNIRASFGSVNPATPTADTYLQ	88						
RESULT 3									
ABP73020									
ID	ABP73020 standard; peptide; 154 AA.								
XX	AC AC								
XX	ABP73020;								
XX	DT DT								
XX	03-JUN-2003 (first entry)								
DE	Amino acid sequence of the ManA carbohydrate binding domain type III.								
KW	ManA; mannanase A; glycoside hydrolase; enzyme; hemicellulose; sugar;								
KW	food; feed; paper pulp; biofuel; mannanase.								
XX	Acidothermus cellulolyticus.								
OS	WO2003012110-A1.								
PN	13-FEB-2003.								
PD	28-JUL-2001; 2001WO-US023819.								
XX	28-JUL-2001; 2001WO-US023819.								
XX	(WIDE ) MIDWEST RES INST.								
PA	Ding S, Adney WS, Vinzant TB, Himmel ME;								
XX	WPI; 2003-248182/24.								
XX	Novel thermal tolerant mannanase A polypeptide derived from Acidothermus								
PT	cellulolyticus, useful for reducing hemicellulose in a starting material,								
PT	for processing of food, and as bulking agents in food stuffs.								
XX	Claim 8; Page 7; 46pp; English.								
PS	The present sequence represents a fragment of ManA, a thermostable								
XX	mannanase A polypeptide derived from Acidothermus cellulolyticus. ManA is								
CC	a member of the glycoside hydrolase family of enzymes. ManA is useful for								
CC	reducing hemicellulose in a starting material to simpler carbohydrate								
CC	units, and ultimately to sugars which are useful in the food, feed, paper								
CC	pulp, and biofuels industries. It is useful for the processing of food								
CC	and in food stuffs as bulking agents, and for the degradation of mannanase.								
CC	ManA is also useful to raise polyclonal and monoclonal antibodies that								
CC	are useful in purifying ManA, or detecting ManA polypeptide expression,								
CC	and as well as reagent tools for characterizing the molecular actions of								
CC	ManA polypeptides								
XX	Sequence 154 AA;								
XX	Query Match								
XX	Best Local Similarity 99.8%; Score 469; DB 6; Length								



```

RESULT 4
ADJ38294
ID ADJ38294 standard; protein; 154 AA.
XX AC
XX ADJ38294;
XX DT
XX 06-MAY-2004 (first entry)
XX DE
XX A. cellulolyticus mannanase Mana carbohydrate binding domain III.
XX KW
XX mannanase A; catalytic domain GH5; carbohydrate binding domain III;
XX KW carbohydrate binding domain II; food processing; foodstuff;
XX KW bulking agent; Mana; enzyme; EC3.2.1.78.
XX OS
XX Acidotherrus cellulolyticus.
XX PN
XX US2003119093-A1.
XX PD
XX 26-JUN-2003.
XX PF
XX 28-JUL-2001; 2001US-00917378.
XX PR
XX 28-JUL-2001; 2001US-00917378.
XX PS
XX (DING/) DING S.
XX PA
XX (ADNE/) ADNEY W S.
XX PA (VINZ/) VINZANT T B.
XX PA (HIMM/) HIMMEL M E.
XX PI
XX Ding S, Adney WS, Vinzant TB, Himmel ME;
XX WPI; 2003-863435/80.
XX DR
XX New composition comprising a purified mannanase A peptide, comprising a
XX PT catalytic domain GH5, carbohydrate binding domain III or carbohydrate
XX PT binding domain II, useful in food processing and in food stuffs as
XX PT bulking agents.
XX PS
XX Claim 7; SEQ ID NO 4; 18pp; English.
XX CC
XX The invention relates to a composition comprising a purified mannanase A
XX CC peptide, having a catalytic domain GH5, carbohydrate binding domain III
XX CC or carbohydrate binding domain II. The composition is useful in food
XX CC processing and in foodstuffs as bulking agents. The present sequence
XX CC represents the amino acid sequence of A. cellulolyticus mannanase Mana
XX CC carbohydrate binding domain III.
XX SQ
XX Sequence 154 AA;
XX
XX Query Match 99.8%; Score 469; DB 7; Length 154;
XX Best Local Similarity 100.0%; Pred. No. 1.1e-45;
XX Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 VSGGVKQVKNDSAPGDNQIKPGLQVNTGSSVDLSTVTYVYFTRDGGSTLVNCD 60
XX Db |||||
XX 1 VSGGVKQVKNDSAPGDNQIKPGLQVNTGSSVDLSTVTYVYFTRDGGSTLVNCD 60
XX
XX QY 61 WAAMGCGNIRASFGSVNPTATPTADTYLQ 88
XX Db |||||
XX 61 WAAMGCGNIRASFGSVNPTATPTADTYLQ 88
XX
XX RESULT 5
XX ABP73022
XX ID ABP73022 standard; protein; 762 AA.
XX AC
XX ABP73022;
XX DT
XX 03-JUN-2003 (first entry)
XX XX
XX Amino acid sequence of the Mana polypeptide.
XX DE
XX Mana; mannanase A; glycoside hydrolase; enzyme; hemicellulose; sugar;

```

```

KW food; feed; paper pulp; biofuel; mannanase.
XX Acidotherrus cellulolyticus.
XX PN
XX WO2003012110-A1.
XX PD
XX 13-FEB-2003.
XX PF
XX 28-JUL-2001; 2001WO-US023819.
XX PR
XX 28-JUL-2001; 2001WO-US023819.
XX KW
XX (MIDE ) MIDWEST RES INST.
XX PI
XX Ding S, Adney WS, Vinzant TB, Himmel ME;
XX WPI; 2003-248182/24.
XX DR
XX N-PSDB; ABZ77633.
XX PT
XX Novel thermal tolerant mannanase A polypeptide derived from Acidotherrus
XX PT cellulolyticus, useful for reducing hemicellulose in a starting material,
XX PT for processing of food, and as bulking agents in food stuffs.
XX PS
XX Claim 3; Page 18-19; 46pp; English.
XX CC
XX The present sequence represents Mana, a thermostable mannanase A
XX CC polypeptide derived from Acidotherrus cellulolyticus. Mana is a member of
XX CC the glycoside hydrolase family of enzymes. Mana is useful for reducing
XX CC hemicellulose in a starting material to simpler carbohydrate units, and
XX CC ultimately to sugars which are useful in the food, feed, paper pulp, and
XX CC biofuels industries. It is useful for the processing of food and in food
XX CC stuffs as bulking agents, and for the degradation of mannanase. Mana is
XX CC also useful to raise polyclonal and monoclonal antibodies that are useful
XX CC in purifying Mana, or detecting Mana polypeptide expression, and as well
XX CC as reagent tools for characterizing the molecular actions of Mana
XX CC polypeptides
XX SQ
XX Sequence 762 AA;
XX
XX Query Match 99.8%; Score 469; DB 6; Length 762;
XX Best Local Similarity 100.0%; Pred. No. 7.8e-45;
XX Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 VSGGVKQVKNDSAPGDNQIKPGLQVNTGSSVDLSTVTYVYFTRDGGSTLVNCD 60
XX Db |||||
XX 455 VSGGVKQVKNDSAPGDNQIKPGLQVNTGSSVDLSTVTYVYFTRDGGSTLVNCD 514
XX
XX QY 61 WAAMGCGNIRASFGSVNPTATPTADTYLQ 88
XX Db |||||
XX 515 WAAMGCGNIRASFGSVNPTATPTADTYLQ 542
XX
XX RESULT 6
XX ADJ38291
XX ID ADJ38291 standard; protein; 762 AA.
XX AC
XX ADJ38291;
XX DT
XX 06-MAY-2004 (first entry)
XX DE
XX A. cellulolyticus mannanase Mana.
XX KW
XX mannanase A; catalytic domain GH5; carbohydrate binding domain III;
XX KW carbohydrate binding domain II; food processing; foodstuff;
XX KW bulking agent; Mana; enzyme; EC3.2.1.78.
XX OS
XX Acidotherrus cellulolyticus.
XX PN
XX US2003119093-A1.
XX PD
XX 26-JUN-2003.
XX PF
XX 28-JUL-2001; 2001US-00917378.

```



XX 28-JUL-2001; 2001US-00917378.  
 PR (DING/) DING S.  
 XX (ADNE/) ADNEY W S.  
 PA (VINZ/) VINZANT T B.  
 PA (HIMM/) HIMMEL M E.  
 XX  
 PI Ding S, Adney WS, Vinzant TB, Himmel ME;  
 XX WPI; 2003-863435/80.  
 DR N-PSDB; ADJ38292.  
 XX  
 PT New composition comprising a purified mannanase A peptide, comprising a  
 PT catalytic domain GH5, carbohydrate binding domain III or carbohydrate  
 PT binding domain II, useful in food processing and in food stuffs as  
 PT bulking agents.  
 XX  
 PS Claim 10; SEQ ID NO 1; 18pp; English.  
 PS The invention relates to a composition comprising a purified mannanase A  
 CC peptide, having a catalytic domain GH5, carbohydrate binding domain III  
 CC or carbohydrate binding domain II. The composition is useful in food  
 CC processing and in foodstuffs as bulking agents. The present sequence  
 CC represents the amino acid sequence of A. cellulolyticus mannanase Mana.  
 XX  
 SQ Sequence 762 AA;  
 Query Match 99.8%; Score 469; DB 7; Length 762;  
 Best Local Similarity 100.0%; Pred. No. 7.8e-45;  
 Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VSGGVKQVQKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYVWFTRDGGSSTLVNCD 60  
 DB 455 VSGGVKQVQKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYVWFTRDGGSSTLVNCD 514  
 QY 61 WAAMCGGNIRASFGSVNPATPTADTYLQ 88  
 DB 515 WAAMCGGNIRASFGSVNPATPTADTYLQ 542  
 RESULT 7  
 ABP73025  
 ID ABP73025 standard; peptide; 150 AA.  
 XX ABP73025;  
 AC  
 XX  
 DT 03-JUN-2003 (first entry)  
 XX  
 DE Amino acid sequence of the GuxA carbohydrate binding domain type III.  
 XX GuxA; cellulase; glycoside hydrolase; enzyme; cellulose; sugar; biofuel;  
 KW detergent; pulp processing; paper processing; feed processing; textile.  
 XX  
 OS Acidotherrmus cellulolyticus.  
 XX  
 PN WO2003012109-A1.  
 XX  
 PD 13-FEB-2003.  
 XX  
 XX 28-JUL-2001; 2001WO-US023817.  
 PF  
 XX 28-JUL-2001; 2001WO-US023817.  
 PR  
 XX (MIDE ) MIDWEST RES INST.  
 PA  
 XX Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;  
 PI WPI; 2003-239526/23.  
 DR  
 XX Novel thermal tolerant GuxA polypeptide derived from Acidotherrmus  
 PT cellulolyticus, useful for reducing cellulose in a starting material, and  
 PT for the conversion of biomass to biofuels and biofuel additives.

XX Claim 6; Page 7; 47pp; English.  
 PS  
 CC ABP73023-28 represent peptide fragments of a GuxA polypeptide. GuxA is a  
 CC thermostable cellulase, and is a member of the glycoside hydrolase family  
 CC of enzymes. GuxA is useful for reducing cellulose in a starting material  
 CC such as agricultural biomass to sugars. This is useful in biofuel  
 CC production. GuxA is also useful in the conversion of biomass to biofuels  
 CC and biofuel additives, in detergents, pulp and paper processing, food and  
 CC feed processing, and in textile process. GuxA is also useful for raising  
 CC polyclonal and monoclonal antibodies that are useful in purifying GuxA,  
 CC or detecting GuxA polypeptide expression, as well as reagent tools for  
 CC characterizing the molecular actions of GuxA polypeptides  
 XX  
 SQ Sequence 150 AA;  
 Query Match 99.1%; Score 466; DB 6; Length 150;  
 Best Local Similarity 98.9%; Pred. No. 2.3e-45;  
 Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VSGGVKQVQKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYVWFTRDGGSSTLVNCD 60  
 DB 1 VSGGVKQVQKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYVWFTRDGGSSTLVNCD 60  
 QY 61 WAAMCGGNIRASFGSVNPATPTADTYLQ 88  
 DB 61 WAAMCGGNIRASFGSVNPATPTADTYLQ 88  
 RESULT 8  
 ADH36640  
 ID ADH36640 standard; protein; 150 AA.  
 XX ADH36640;  
 AC  
 XX 11-MAR-2004 (first entry)  
 DT  
 XX A. cellulolyticus glycoside hydrolase, GuxA CBD type III domain.  
 DE Thermal tolerant cellulase; glycoside hydrolase; GuxA;  
 KW carbohydrate degradation; cellulose; agricultural biomass;  
 KW municipal solid waste; thermostable; enzyme;  
 KW carbohydrate binding domain type III; CBD III.  
 XX  
 OS Acidotherrmus cellulolyticus.  
 XX  
 PN US2003104522-A1.  
 XX  
 PD 05-JUN-2003.  
 XX  
 XX 28-JUL-2001; 2001US-00917383.  
 PF  
 XX 28-JUL-2001; 2001US-00917383.  
 PR  
 XX (DING/) DING S.  
 PA (ADNE/) ADNEY W S.  
 PA (VINZ/) VINZANT T B.  
 PA (HIMM/) HIMMEL M E.  
 PA (DECK/) DECKER S R.  
 XX  
 PI Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;  
 XX WPI; 2004-106451/11.  
 DR  
 XX Novel isolated thermostable GuxA polypeptide useful for detecting  
 PT polynucleotide encoding GuxA, assessing carbohydrate degradation activity  
 PT of GuxA, reducing cellulose in starting material e.g., agricultural  
 PT biomass.  
 XX  
 PS Claim 27; SEQ ID NO 5; 20pp; English.  
 XX The present invention relates to the isolation of a thermal tolerant  
 CC cellulase from Acidotherrmus cellulolyticus. The cellulase is a member of



CC the glycoside hydrolase family and is designated GuxA. Also disclosed are  
 CC methods of producing and using GuxA. The GuxA polypeptide is useful for  
 CC the detection of a polynucleotide encoding GuxA. The polypeptide sequence  
 CC is also useful for assessing the carbohydrate, e.g. cellulose,  
 CC degradation activity of GuxA. GuxA is useful for reducing cellulose in a  
 CC starting material such as agricultural biomass or municipal solid waste.  
 CC The polypeptide molecule of GuxA is thermostable. The present sequence  
 CC represents a domain of the A. cellulolyticus GuxA polypeptide.  
 XX  
 SQ Sequence 150 AA;

Query Match 99.1%; Score 466; DB 8; Length 150;  
 Best Local Similarity 98.9%; Pred. No. 2.3e-45;  
 Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VSGGVKQVYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYVWFTRDGSGSTLVYVNC 60  
 DB 1 VSGGLKVQYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYVWFTRDGSGSTLVYVNC 60  
 QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
 DB 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88

RESULT 9  
 ABP73029  
 ID ABP73029 standard; protein; 1228 AA.

AC ABP73029;  
 DT 03-JUN-2003 (first entry)

DE Amino acid sequence of the GuxA polypeptide.

KW GuxA; cellulase; glycoside hydrolase; enzyme; cellulose; sugar; biofuel;  
 KW detergent; pulp processing; paper processing; feed processing; textile.

OS Acidothermus cellulolyticus.  
 XX W02003012109-A1.  
 XX 13-FEB-2003.

XX 28-JUL-2001; 2001WO-US023817.

XX 28-JUL-2001; 2001WO-US023817.

XX (MIDE ) MIDWEST RES INST.

XX Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;

XX WPI; 2003-239526/23.

XX N-PSDB; ABZ77634.

XX Novel thermal tolerant GuxA polypeptide derived from Acidothermus  
 PT cellulolyticus, useful for reducing cellulose in a starting material, and  
 PT for the conversion of biomass to biofuels and biofuel additives.

XX Claim 3; Page 19; 47pp; English.

XX The present sequence represents a GuxA polypeptide. GuxA is thermostable  
 CC cellulase, and is a member of the glycoside hydrolase family of enzymes.  
 CC GuxA is useful for reducing cellulose in a starting material such as  
 CC agricultural biomass to sugars. This is useful in biofuel production.  
 CC GuxA is also useful in the conversion of biomass to biofuels and biofuel  
 CC additives, in detergents, pulp and paper processing, food and feed  
 CC processing, and in textile process. GuxA is also useful for raising  
 CC polyclonal and monoclonal antibodies that are useful in purifying GuxA,  
 CC or detecting GuxA polypeptide expression, as well as reagent tools for  
 CC characterizing the molecular actions of GuxA polypeptides

XX Sequence 1228 AA;

Query Match 99.1%; Score 466; DB 6; Length 1228;  
 Best Local Similarity 98.9%; Pred. No. 3.1e-44;  
 Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VSGGVKQVYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYVWFTRDGSGSTLVYVNC 60  
 DB 584 VSGGLKVQYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYVWFTRDGSGSTLVYVNC 643  
 QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
 DB 644 WAAMCGNIRASFGSVNPATPTADTYLQ 671

RESULT 10  
 ADH36636

ID ADH36636 standard; protein; 1228 AA.

XX ADH36636;

XX 11-MAR-2004 (first entry)

XX Acidothermus cellulolyticus glycoside hydrolase, GuxA.

KW Thermal tolerant cellulase; glycoside hydrolase; GuxA;  
 KW carbohydrate degradation; cellulose; agricultural biomass;  
 KW municipal solid waste; thermostable; enzyme.

XX Acidothermus cellulolyticus.

XX US2003104522-A1.

XX 05-JUN-2003.

XX 28-JUL-2001; 2001US-00917383.

XX 28-JUL-2001; 2001US-00917383.

XX (DING/) DING S.

XX (ADNE/) ADNEY W S.

XX (VINZ/) VINZANT T B.

XX (HIMM/) HIMMEL M E.

XX (DECK/) DECKER S R.

XX Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;

XX WPI; 2004-106451/11.

XX N-PSDB; ADH36637.

XX Novel isolated thermostable GuxA polypeptide useful for detecting  
 PT polynucleotide encoding GuxA, assessing carbohydrate degradation activity  
 PT of GuxA, reducing cellulose in starting material e.g., agricultural  
 PT biomass.

XX Claim 27; SEQ ID NO 1; 20pp; English.

XX The present invention relates to the isolation of a thermal tolerant  
 CC cellulase from Acidothermus cellulolyticus. The cellulase is a member of  
 CC the glycoside hydrolase family and is designated GuxA. Also disclosed are  
 CC methods of producing and using GuxA. The GuxA polypeptide is useful for  
 CC the detection of a polynucleotide encoding GuxA. The polypeptide sequence  
 CC is also useful for assessing the carbohydrate, e.g. cellulose,  
 CC degradation activity of GuxA. GuxA is useful for reducing cellulose in a  
 CC starting material such as agricultural biomass or municipal solid waste.  
 CC The polypeptide molecule of GuxA is thermostable. The present sequence  
 CC represents A. cellulolyticus glycoside hydrolase, GuxA.

XX Sequence 1228 AA;

Query Match 99.1%; Score 466; DB 8; Length 1228;  
 Best Local Similarity 98.9%; Pred. No. 3.1e-44;  
 Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYVWFTRDGSGSTLVYVNC 60







KW Avicelase; Aviii; glycoside hydrolase; enzyme; cellulase; biofuel;  
KW detergent; pulp processing; paper processing; feed processing; textile;  
KW cellulose.

OS *Acidothermus cellulolyticus*.

XX	Key	Location/Qualifiers
FH		

```

FT      Misc-difference 89
FT      /note= "unspecified residue"

```

AA PN WO2003012090-A2.

13-FFB-2003.

28-JUL-2001: 2001WO-US023818.

28-JUL-2001: 2001WO-US023818.

XX  
PA (MIDE ) MIDWEST RES INST.

XX  
PT Ding S. Adnev WS. Vinzant TB. Himmel ME:

XX  
DB  
WPT: 2003-248177/24.

xx PT New thermostable AvIII peptide from *Acidothermus cellulolyticus*, useful  
PT for degradation of cellulose or in generating anti-AvIII antibodies for  
PT purifying recombinant AvIII polypeptides from genetically engineered  
PT host cells.

XX  
PS  
Claim 6: Page 8

OS *Acidothermus cellulolyticus*.

XX	FH	Key	Location/Qualifiers

FT	Misc-difference 89
FT	/note= "A string of unknown amino acids (50 or more)"

XX PN US2004038334-A1.

XX  
PD  
26-FEB-2004.

28-JUL-2001: 2001US-00917376.

28-JUL-2001: 2001US-00917376.

PA (DING/) DING S.

PA (ADNE//) ADNEY W S.

PA (VINZ/) VINZANT T B.

PA (HIMM/) HIMMEL M E.

XX  
PI  
ning S  
Adney WS  
Vinzant TR  
Himmel ME:

XX  
WPB 2004-203224/19

XX  
PT  
Novel thermostable XVIIII polyimide

PT isolated from *Acidothermus cellulolyticus*, useful for degrading treated  
XX biomass into simpler forms of carbohydrate.  
XX  
XX Claim 28; SEQ ID NO 4; 19pp; English.  
XX  
XX The invention relates to a thermostable cellulase enzyme, avicelase III  
CC (AviIII) and its nucleic acid sequence. AviIII is useful for reducing  
CC cellulose in a starting material. A thermostable AviIII peptide is useful  
CC for degrading treated biomass into simpler forms of carbohydrate, which  
CC is used in the formation of ethanol or other industrial chemicals. It is  
CC also useful for treating fabrics to remove cellulose-containing stains.  
CC The present sequence is *Acidothermus cellulolyticus* AviIII carbohydrate  
CC binding domain type III (CBD III) partial protein.

Sequence 89 AA:

DE A. cellulolyticus AviiI CBD III partial protein #1.

Key	Location/Qualifiers
-----	---------------------

FT	Misc-difference	957	
FT		/note=	"unspecified residue encoded by N"

XX



PN WO2003012090-A2.  
XX 13-FEB-2003.  
XX 28-JUL-2001; 2001WO-US023818.  
PF XX  
XX 28-JUL-2001; 2001WO-US023818.  
PR XX  
XX (WIDE ) MIDWEST RES INST.  
PA XX  
XX Ding S, Adney WS, Vinzant TB, Himmel ME;  
PI XX  
XX WPI; 2003-248177/24.  
DR XX  
DR N-PSDB; ABZ77632.  
XX XX  
PT New thermostable Aviii peptide from Acidothermus cellulolyticus, useful  
PT for degradation of cellulose or in generating anti-Aviii antibodies for  
PT purifying recombinant Aviii polypeptides from genetically engineered  
PT host cells.  
XX XX  
PS Claim 2; Page 20; 44pp; English.  
PS XX  
CC The present sequence represents a thermostable avicelase polypeptide,  
CC designated Aviii. Aviii is a member of the glycoside hydrolase family  
CC of enzymes, and is a cellulase. Aviii is useful in the conversion of  
CC biomass to biofuels and biofuel additives. It may be useful in the  
CC production of detergents, pulp and paper processing, food and feed  
CC processing and in textile processes. The thermostable Aviii peptide is  
CC useful in the degradation of cellulose, and in generating specific anti-  
CC Aviii antibodies that are useful in purifying recombinant Aviii  
CC polypeptides from genetically engineered host cells, in detecting Aviii  
CC polypeptide expression, as well as a reagent tool for characterizing the  
CC molecular actions of the polypeptide. The Aviii polynucleotide is useful  
CC as a source of probes or primers in various diagnostic assays  
XX XX  
SQ Sequence 957 AA;  
Query Match 98.3%; Score 462; DB 6; Length 957;  
Best Local Similarity 97.7%; Pred. No. 6.6e-44;  
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 VSGGVKQYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRWFTRDGGSTLVNCD 60  
Db 869 VSGGVKQYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYRWFTRDGGSTLVNCD 928  
Qy 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
Db 929 WAAIGCGNIRASFGSVNPATPTADTYLQ 956  
Search completed: October 5, 2005, 08:03:11  
Job time : 18.0496 secs



Result No.	Score	Query Match	Length	DB	ID	Description
1	191	40.6	1711	2	T31337	1,4-beta-glucanase
2	190.5	40.5	1779	2	T31085	xylanase - Caldiche
3	190	40.4	508	2	G69593	cellulase (EC 3.2.
4	190	40.4	508	2	A28874	cellulase (EC 3.2.
5	188	40.0	499	2	JN0111	cellulase (EC 3.2.
6	186.5	39.7	1331	2	A48954	mannan endo-1,4-be
7	186.5	39.7	1742	2	T17120	cellulase (EC 3.2.
8	186	39.6	499	2	A27198	cellulase (EC 3.2.
9	182.5	38.8	915	2	A43802	cellulase (EC 3.2.
10	182.5	38.8	1039	2	S02711	cellulase (EC 3.2.
11	180.5	38.4	145	2	A41897	cellulase homolog
12	180	38.3	486	2	I40548	bifunctional cellul
13	172	36.6	700	2	B41897	cellulase (EC 3.2.
14	165.5	35.2	879	2	A47704	endoglucanase I (E
15	165	35.1	1854	2	S36859	c1pA protein - Clo
16	149.5	31.8	505	2	S39962	endoglucanase - Er
17	147.5	31.4	504	2	S54744	cellulase (EC 3.2.
18	141.5	30.1	986	2	S12021	thermoactive cellul
19	130.5	27.8	586	2	PC6006	scaffolding protei
20	130.5	27.8	1483	2	C97012	probably celluloso
21	126.5	26.9	1162	2	T30433	scaffolding protei
22	117	24.9	1230	2	S47466	cellulose 1,4-beta
23	114	24.3	1848	2	A44140	cellulose-binding
24	75.5	16.1	618	2	T08685	hypothetical prote
25	73.5	15.6	547	2	T25478	hypothetical prote
26	73.5	15.6	1428	2	AC3224	hypothetical prote
27	70	14.9	5188	2	B85547	probable RTX fami
28	68.5	14.9	5291	2	F90966	hypothetical prote
29	68.5	14.6	5574	2	A69196	cell surface glyco

Query Match 40.5%: Score 190.5; DB 2; Length 1779;



Best Local Similarity 43.5%; Pred. No. 1.4e-11;  
Matches 37; Conservative 14; Mismatches 33; Indels 1; Gaps 1;

QY 4 GVKVOYKNDSPAGDNQIKPGLQVNTGSSVDLSVTTVRYWFFTRDGGSSTLVYNCDDAA 63  
Db 1113 GLKLYLNKNNSTASTSRPWFKTVNGGSSVDLSRVKIRYWTVDGDKPOSAY-CDWAQ 1171

QY 64 MGCNIRASFGSVNPATPTADTYLQ 88  
Db 1172 IGASNVTFNFKLTSGVGADTYLE 1196

RESULT 3  
G69593  
cellulase (EC 3.2.1.4) bglC precursor - Bacillus subtilis  
N:Alternate names: endo-1,4-beta-glucanase  
C:Species: Bacillus subtilis  
C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 09-Jul-2004  
C:Accession: G69593; A26114; I40353; S24239; S49103; I39803  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berteaux, C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choudhury, M.; Takakoshi, A.; Tanaka, T.; Terpestra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yanane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
A:Reference number: A69580; MUID:98044033; PMID:9384377  
A:Accession: G69593  
A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA  
A:Residues: 1-508 <KUN>  
A:Cross-references: UNIPROT:P10475; UNIPROT:Q03882; GB:Z99113; GB:AL009126; NID:G2634090  
A:Experimental source: strain 168  
R:MacKay, R.M.; Lo, A.; Willick, G.; Zuker, M.; Baird, S.; Dove, M.; Moranelli, F.; Selig  
Nucleic Acids Res. 14, 9159-9170, 1986  
A:Title: Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene.  
A:Reference number: A26114; MUID:87066783; PMID:3024130  
A:Accession: A26114  
A:Molecule type: DNA  
A:Residues: 10-508 <MAC>  
A:Cross-references: GB:X04689; NID:G39823; PIDN:CAA28392.1; PID:G39824  
A:Experimental source: strain PAP115  
A:Note: part of this sequence, including the amino end of the mature form, was confirmed  
R:Lindahl, V.; Aa, K.; Tronemo, A.  
Antonie Van Leeuwenhoek 66, 327-332, 1994  
A:Title: Nucleotide sequence of an endo-beta-1,4-glucanase gene from Bacillus subtilis C  
A:Reference number: I40353; MUID:95225656; PMID:7710280  
A:Accession: I40353  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 10-291, 'N', 293-508 <LIN1>  
A:Cross-references: EMBL:X67044; NID:G39776; PIDN:CAA47429.1; PID:G39777  
R:Lindahl, V.; Aa, K.  
submitted to the EMBL Data Library, June 1992  
A:Reference number: S24239  
A:Accession: S24239  
A:Molecule type: DNA  
A:Residues: 10-291, 'N', 293-508 <LIN2>  
A:Cross-references: EMBL:X67044; NID:G39776; PIDN:CAA47429.1; PID:G39777  
R:Wolf, M.; Geczi, A.; Borris, R.  
submitted to the EMBL Data Library, December 1993  
A:Description: Genes encoding beta glucan-hydrolyzing enzymes in Bacillus subtilis: cons  
A:Reference number: S49103  
A:Accession: S49103

A:Molecule type: DNA  
A:Residues: 10-508 <WOL>  
A:Cross-references: EMBL:Z29076; NID:G509266; PIDN:CAA82317.1; PID:G509267  
R:Seo, Y.S.; Lee, Y.H.; Pek, U.H.; Kang, H.  
Korean J. Microbiol. 24, 236-242, 1986  
A:Title: Analysis on the nucleotide sequence of the signal region of Bacillus subtilis  
A:Reference number: I39803  
A:Accession: I39803  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 10-13, 'V', 15, 'N', 17-18, 'V', 20-21, 'F', 23, 'A', 25-26, 'AI', 29-31, 'P', 33, 'PQ', 36  
A:Cross-references: GB:M38634; NID:G142657; PIDN:AAA22300.1; PID:G142658  
A:Experimental source: strain ATCC 6633  
C:Comment: The low molecular weight of the mature protein suggests carboxyl-terminal pro  
C:Genetics:  
A:Gene: bglC  
C:Function:  
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel  
A:Pathway: cellulose degradation  
C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation  
F:1-38/Domain: signal sequence #status predicted <SIG>  
F:39-508/Product: cellulase #status predicted <MAT>

Query Match 40.4%; Score 190; DB 2; Length 508;  
Best Local Similarity 40.7%; Pred. No. 4.2e-12;  
Matches 35; Conservative 21; Mismatches 26; Indels 4; Gaps 2;

QY 4 GVKVOYKNDSPAGDNQIKPGLQVNTGSSVDLSVTTVRYWFFTRDGGSSTLVYNCDDAA 62  
Db 362 GISVQYRAGDSMNSNQIRPQLIKNGNNTVDLKDVTARYWYKAKNGQN---FDCDYA 418

QY 63 AMGCCNIRASFGSVNPATPTADTYLQ 88  
Db 419 QIGCGNVTHKFTVTLHKPKQADTYLE 444

RESULT 4  
A26874  
cellulase (EC 3.2.1.4) precursor - Bacillus subtilis (strain DLG)  
N:Alternate names: endo-1,4-beta-glucanase  
C:Species: Bacillus subtilis  
C>Date: 08-Mar-1989 #sequence\_revision 08-Mar-1989 #text\_change 09-Jul-2004  
A:Accession: A26874; B26874  
R:Robson, L.M.; Chambliss, G.H.  
J. Bacteriol. 169, 2017-2025, 1987  
A:Title: Endo-beta-1,4-glucanase gene of Bacillus subtilis DLG.  
A:Reference number: A26874; MUID:87194581; PMID:3106328  
A:Accession: A26874  
A:Molecule type: DNA  
A:Residues: 1-508 <ROB1>  
A:Cross-references: UNIPROT:P07983; GB:M16185; NID:G143007; PIDN:AAA22496.1; PID:G143008  
A:Experimental source: strain DLG  
A:Accession: B26874  
A:Molecule type: protein  
A:Residues: 39-53 <ROB2>  
A:Experimental source: strain DLG  
A:Note: the authors believe Met-1 and Met-2 may be alternate initiators  
C:Comment: The low molecular weight of the mature protein suggests carboxyl-terminal pro  
C:Function:  
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel  
A:Pathway: cellulose degradation  
C:Keywords: alternative initiators; extracellular protein; glycosidase; hydrolase; poly  
F:1-38/Domain: (or 2-38) signal sequence #status predicted <SIG>

Query Match 40.4%; Score 190; DB 2; Length 508;  
Best Local Similarity 43.0%; Pred. No. 4.2e-12;  
Matches 37; Conservative 17; Mismatches 28; Indels 4; Gaps 2;

QY 4 GVKVOYKNDSPAGDNQIKPGLQVNTGSSVDLSVTTVRYWFFTRDGGSSTLVYNCDDAA 62  
Db 362 GVSVOYKAGDGRVNSNQIRPQLIKNGNNTVDLKDVTARYWYKAKNGQN---FDCDYA 418

QY 63 AMGCCNIRASFGSVNPATPTADTYLQ 88











Db 411 LGCNVSHVTYTLHKPKGADTYLE 435

RESULT 13

B41897

Cellulase (EC 3.2.1.4) - Bacillus lautus

N;Alternate names: endo-1,4-beta-glucanase

C;Species: Clostridium thermoecellum

C;Date: 10-Mar-1994 #sequence\_revision 10-Mar-1994 #text\_change 09-Jul-2004

C;Accession: B41897; S27499

R;Hansen, C.K.; Diderichsen, B.; Jorgensen, P.L.

J. Bacteriol. 174, 3522-3531, 1992

A;Title: celsa from Bacillus lautus PL236 encodes a novel cellulose-binding endo-beta-1,4

A;Reference number: A41897; MUID:92276330; PMID:1592807

A;Accession: B41897

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-700 <HAN>

A;Cross-references: UNIPROT:P29719; EMBL:M76588; NID:g142661; PIDN:AAA22303.1; PID:g142661

A;Experimental source: PL236

A;Note: sequence extracted from NCBI backbone (NCBIP:104605)

C;Function:

A;Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel

A;Pathway: cellulose degradation

C;Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 36.6%; Score 172; DB 2; Length 700;

Best Local Similarity 44.9%; Pred. No. 4.4e-10;

Matches 40; Conservative 18; Mismatches 27; Indels 4; Gaps 3;

QY 1 VSGGVKVOYKND-SAPGDNQIKPGLQLVNTGSSVDLSTVTYVWFTRDGSSSTLVYNC 59

Db 549 VNSDLVQYKGDNRNATNDQIKPHFNQKSPDLSLTLRYFTKD-SSAAMNGWI 607

QY 60 DWAMGCGNIRASFGSVNPATPTADTYLQ 88

Db 608 DWAKLGSGNIQISFGNHGA--DSDTYAB 634

RESULT 14

A47704

endoglucanase I (EC 3.2.1.-) Cell - Clostridium thermoecellum

C;Species: Clostridium thermoecellum

C;Date: 19-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004

C;Accession: A47704

R;Hawlewood, G.P.; Davidson, K.; Laurie, J.I.; Huskisson, N.S.; Gilbert, H.J.

J. Gen. Microbiol. 139, 307-316, 1993

A;Title: Gene sequence and properties of CellI, a family E endoglucanase from Clostridium

A;Reference number: A47704; MUID:93171873; PMID:8436949

A;Accession: A47704

A;Status: preliminary

A;Molecule type: DNA; protein

A;Residues: 1-879 <HAZ>

A;Cross-references: UNIPROT:Q02934; GB:I04735; NID:g144807; PIDN:AAA20892.1; PID:g144808

A;Note: sequence extracted from NCBI backbone (NCBIN:125637, NCBIP:125638)

C;Keywords: glycosidase; hydrolase

Query Match 35.2%; Score 165.5; DB 2; Length 879;

Best Local Similarity 33.0%; Pred. No. 2.7e-09;

Matches 29; Conservative 26; Mismatches 32; Indels 1; Gaps 1;

QY 1 VSGGVKVOYKNDSPGDNQIKPGLQLVNTGSSVDLSTVTYVWFTRDGSSSTLVYNC 60

Db 736 IKGEVLQYANGAGTNSINPRKLIINNGKAINLSVDKIRYYTYKGGASQNF-CD 794

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88

Db 795 WSSAGNSNVGTGFFNLSPKSGADTCLE 822

RESULT 15

S36859

cipA protein - Clostridium thermoecellum

N;Alternate names: probable cellulosome protein large chain SL

C;Species: Clostridium thermoecellum

C;Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 09-Jul-2004

C;Accession: S36859; S33527; S25767; S28659; T18261

R;Gerengross, U.T.; Demain, A.L.

submitted to the EMBL Data Library, January 1993

A;Reference number: S36859

A;Accession: S36859

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1854 <GER>

A;Cross-references: UNIPROT:Q06851; EMBL:L08665

R;Gerengross, U.T.; Romaniec, M.P.M.; Kobayashi, T.; Huskisson, N.S.; Demain, A.L.

Mol. Microbiol. 8, 325-334, 1993

A;Title: Sequencing of a Clostridium thermoecellum gene (cipA) encoding the cellulosomal

A;Reference number: S33527; MUID:93302508; PMID:8316083

A;Accession: S33527

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1615,1617-1854 <GE2>

A;Cross-references: EMBL:L08665

R;Fujino, T.; Beguin, P.; Aubert, J.P.

FMMS Microbiol. Lett. 94, 165-170, 1992

A;Title: Cloning of a Clostridium thermoecellum DNA fragment encoding polypeptides that

A;Reference number: S25767

A;Accession: S25767

A;Molecule type: DNA

A;Residues: 1216-1232,'Y',1234-1241,'T',1243-1319,'R',1321-1615,1617-1854 <FU2>

A;Cross-references: EMBL:X67406

R;Beguin, P.

submitted to the EMBL Data Library, August 1992

A;Reference number: S28659

A;Accession: S28659

A;Molecule type: DNA

A;Residues: 1216-1232,'Y',1234-1241,'T',1243-1319,'R',1321-1615,1617-1768,'R',1770-1854

R;Fujino, T.; Beguin, P.; Aubert, J.P.

J. Bacteriol. 175, 1891-1899, 1993

A;Title: Organization of a Clostridium thermoecellum gene cluster encoding the celluloso

e.

A;Reference number: Z18847; MUID:93209931; PMID:8458832

A;Accession: T18261

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1821-1854 <FU2>

A;Cross-references: EMBL:X67506; NID:g296879; PID:g296880; PIDN:CAA47840.1

C;Genetics:

A;Gene: cipA

Query Match 35.1%; Score 165; DB 2; Length 1854;

Best Local Similarity 36.1%; Pred. No. 6.7e-09;

Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKVOYKNDSPGDNQIKPGLQLVNTGSSVDLSTVTYVWFTRDGSSSTLVYNC 60

Db 365 VSGNLKVEFYNSPSTTNSINPQKVTYNTGSSAIDLSKLTLYRYTYVDGQKDTFW-CD 423

QY 61 WAAM-----GCGNIRASFGSVNPATPTADTYLQ 88

Db 424 HAAIIGSGSYNGITSNVKGTFVKMSSSTNNADTYLE 460

Search completed: October 5, 2005, 07:57:09

Job time : 3.46958 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 07:44:57 ; Search time 11.0656 Seconds  
(without alignments)  
4118.611 Million cell updates/sec

Title: US-09-917-376-4

Perfect score: 470

Sequence: 1 VSGVKVQYKNDSPGDNQ.....RASFGSVNPATPTADTVLQX 89

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_03.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	257	54.7	741	2 Q82QF2	Q82qf2 streptomyce
2	228.5	48.6	616	2 Q7X2N2	Q7x2n2 thermomono
3	225.5	48.0	170	2 Q9RFK6	Q9rfx6 caldibacill
4	216.5	46.1	930	2 Q9RFX5	Q9rfx5 caldibacill
5	214.5	45.6	921	2 Q9L8L8	Q9l8l8 caldibacill
6	209.5	44.6	1091	2 Q8KKF7	Q8kkf7 paenibacill
7	204.5	43.5	997	2 Q9Z4I1	Q9z4i1 bacillus sp
8	194.5	41.4	1751	2 Q9AQG4	Q9aqg4 caldicellul
9	192.5	41.0	1000	2 Q24820	O24820 thermophil
10	192.5	41.0	1770	2 Q9X3P5	Q9x3p5 caldicellul
11	191.5	40.7	261	2 Q9AQG7	Q9aqg7 caldicellul
12	191.5	40.7	1426	2 Q9X3P6	Q9x3p6 caldicellul
13	191	40.6	1711	2 P96311	P96311 anaerocellu
14	190.5	40.5	996	2 Q9AQH0	Q9aqh0 caldicellul
15	190.5	40.5	1779	2 Q52374	O52374 caldicellul
16	190	40.4	499	1 GUN1_BACSU	P07983 bacillus su
17	190	40.4	499	1 GUN2_BACSU	P10475 bacillus su
18	190	40.4	499	2 Q93TJ6	Q93tj6 bacillus su
19	190	40.4	508	2 Q93LD0	Q93ld0 bacillus su
20	190	40.4	508	2 Q70K05	Q70k05 bacillus am
21	188	40.0	499	1 GUN3_BACSU	P23549 bacillus su
22	187	39.8	499	2 Q52731	O52731 bacillus sp
23	187	39.8	499	2 Q674Q1	Q6y4q1 bacillus su
24	186.5	39.7	1331	1 MANB_CALSA	P22533 caldocellum
25	186.5	39.7	1742	1 GUNA_CALSA	P22534 caldocellum
26	186	39.6	499	2 Q45532	Q45532 bacillus su
27	186	39.6	501	2 Q83012	Q83012 bacillus sp
28	182.5	38.8	1039	1 GUNB_CALSA	P10474 c endogluc
29	180.5	38.4	145	1 YCEA_PAEIA	P29718 paenibacill
30	180	38.3	486	2 Q45430	Q45430 bacillus sp
31	177	37.7	499	2 Q8RPQ6	Q8rpq6 bacillus am

32 172.5 36.7 1915 2 Q9RPL0 acetivibrio  
33 172 36.6 700 1 GUNA\_PAEIA P29719 paenibacill  
34 168 35.7 473 2 Q9RK75 Q9rk75 streptomyce  
35 165.5 35.2 879 1 GUN1\_CLOTM Q02934 clostridium  
36 165.5 35.2 887 2 Q9L3J8 Q9l3j8 clostridium  
37 165 35.1 772 1 CIEB\_CLOTM Q01866 clostridium  
38 165 35.1 1853 2 CIEB\_CLOTM Q06851 clostridium  
39 157 33.4 534 2 Q65IY7 Q65iy7 bacillus li  
40 157 33.4 542 2 Q7X386 Q7x386 bacillus li  
41 150.5 32.0 2319 2 Q9FDJ9 Q9fdj9 bacteroides  
42 149.5 31.8 444 1 GUNN\_ERWCA O59394 erwinia car  
43 149.5 31.8 505 1 GUNW\_ERWCA O47096 erwinia car  
44 147.5 31.4 504 1 GUNW\_ERWCA Q59395 erwinia car  
45 147.5 31.4 914 1 GUX2\_CLOSR P50900 clostridium

ALIGNMENTS

RESULT 1

Q82QF2 ID Q82QF2 PRELIMINARY; PRT; 741 AA.  
AC Q82QF2;  
DT 01-JUN-2003 (Tremblrel. 24, Created)  
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
DE Putative cellulose 1,4-beta-cellobiosidase.  
GN Name=guxal; OrderedLocustNames=SAV557;  
OS Streptomyces avermitilis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=33903;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MA-4680;  
RX MEDLINE=21477403; PubMed=11572948; DOI=10.1073/pnas.211433198;  
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,  
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonce T.,  
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.,  
RT "Genome sequence of an industrial microorganism Streptomyces  
RT avermitilis; deducing the ability of producing secondary  
RT metabolites.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MA-4680;  
RX MEDLINE=22608306; PubMed=12692562;  
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,  
RA Sakaki Y., Hattori M., Omura S.,  
RT "Complete genome sequence and comparative analysis of the industrial  
RT microorganism Streptomyces avermitilis.";  
RL Nat. Biotechnol. 21:526-531(2003).  
CC -!- SIMILARITY: Contains 1 fibronectin type III domain.  
DR EMBL; AP005023; BAC68267.1; -.  
DR HSSP; P20533; 1K85  
DR GO; GO:0004553; Phydrolase activity, hydrolyzing O-glycosyl . . .; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR InterPro; IPR001956; CBD\_3.  
DR InterPro; IPR008965; Cellul bind.  
DR InterPro; IPR003961; FN\_III-like.  
DR InterPro; IPR008957; FN\_III-like.  
DR InterPro; IPR01524; Glyco\_hydro\_6.  
DR InterPro; IPR011253; Glyco\_hydro\_6-1k.  
DR Pfam; PF00942; CBM\_3; 1.  
DR DR Pfam; PF00041; fn3; 1.  
DR Pfam; PF01341; Glyco\_hydro\_6; 1.  
DR PRINTS; PR00733; GLHYDRLASE6.  
DR ProDom; PD001947; CBD\_3; 1.  
DR ProDom; PD003733; Glyco\_hydro\_6; 1.  
DR SMART; SM00060; FN3; 1.  
DR PROSITE; PS50853; FN3; 1.  
DR PROSITE; PS00655; GLYCOSYL\_HYDROL\_F6\_1; 1.  
KW Complete proteome.



RA	Sunna A., Gibbs M.D., Chin C.W.J., Nelson P.J., Bergquist P.L.;
RT	"A gene encoding a novel multidomain beta-1,4-mannanase from
RI	Caldibacillus cellulovorans and action of the recombinant enzyme on
RJ	kraft pulp";
RL	Appl. Environ. Microbiol. 66:664-670(2000).
RR	EMBL; AF163837; AAF22273.1; -.
DR	HSSP; Q06851; INBC.
DR	GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR	InterPro; IPR001956; CBD_3.
DR	InterPro; IPR008965; Cellul_bind.
DR	Pfam; PF00942; CBM_3; 1.
DR	ProDom; PD001947; CBD_3; 1.
KW	Hypothetical protein.
FT	NON_TER 1
SQ	SEQUENCE 170 AA; 18493 MW; 7AC9D33F44E3A0B4 CRC64;
Query Match	48.0%; Score 225.5; DB 2; Length 170;
Best Local Similarity	47.7%; Pred. No. 8.3e-17;
Matches	41; Conservative 17; Mismatches 27; Indels 1; Gaps 1;
QY	3 GGVKVOYKNNDSPAGDNQIKPGIQLVNTGSSVDLSTVTVVYWFTRDGGSTLVNCDWA 62   :       :   :         :         :         :         :         20 GSLVVGYRADTWNAGDNLPHFRIVNRGTSSVPLSELTIKYWTVD-GDKPQVFNCDA 78
DB	63 AMCGCINIRASFGSVNPATPTADTYLQ 88 :         :   :     :       :
QY	79 QVGCSNVRGSFVKLSTGRTGADYYIE 104 :         :   :     :       :
DB	
RESULT 4	
Q9RFX5	PRELIMINARY; PRT; 930 AA.
ID	Q9RFX5
AC	Q9RFX5;
DT	01-MAY-2000 (TREMBLrel. 13, Created)
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT	01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE	Multidomain beta-1,4-mannanase precursor.
OS	Name=manA;
GN	Caldibacillus cellulovorans.
OC	Bacteria; Firmicutes; Bacillales; Alicyclobacillaceae; Caldibacillus.
OX	NCBI_TaxID=74586;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=20120520; PubMed=10653733; DOI=10.1128/AEM.66.2.664-670.2000;
RA	Sunna A., Gibbs M.D., Chin C.W.J., Nelson P.J., Bergquist P.L.;
RT	"A gene encoding a novel multidomain beta-1,4-mannanase from
RI	Caldibacillus cellulovorans and action of the recombinant enzyme on
RJ	kraft pulp";
RL	Appl. Environ. Microbiol. 66:664-670(2000).
RR	EMBL; AF163837; AAF22274.1; -.
DR	HSSP; Q9ZF13; 3MAN.
DR	GO; GO:0019028; C:viral capsid; IEA.
DR	GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR	GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR	InterPro; IPR001956; CBD_3.
DR	InterPro; IPR008965; Cellul bind.
DR	InterPro; IPR004302; Chitin binding_3.
DR	InterPro; IPR001547; Glyco_hydro_5.
DR	Pfam; PF00942; CBM_3; 2.
DR	Pfam; PF00150; Cellulase; 1.
DR	Pfam; PF03067; Chitin_bind 3; 1.
DR	ProDom; PD001947; CBD_3; 2.
DR	PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
FT	SIGNAL.
FT	SIGNAL 1 33 Potential.
FT	CHAIN 34 930 multidomain beta-1,4-mannanase.
SQ	SEQUENCE 930 AA; 101576 MW; 0086638D54D1A2CC CRC64;
Query Match	46.1%; Score 216.5; DB 2; Length 930;
Best Local Similarity	46.5%; Pred. No. 5.9e-15;
Matches	40; Conservative 17; Mismatches 28; Indels 1; Gaps 1;
QY	3 GGVKVOYKNNDSPAGDNQIKPGIQLVNTGSSVDLSTVTVVYWFTRDGGSTLVNCDWA 62



OC	Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.
OX	NCBI_TaxID=198119;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=BP-23;
RL	Sanchez M.M., Pastor F.I.J., Diaz P.;
RA	Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC	-1- SIMILARITY: Contains 2 fibronectin type III domains.
DR	EMBL; AJ488933; CAD32945.1; -.
DR	HSSP; P38696; 111Y.
DR	GO; GO:0016162; F:cellulose 1,4-beta-cellobiosidase activity; IEA.
DR	GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR	InterPro; IPR001956; CBD_3.
DR	InterPro; IPR008965; Cellul_bind.
DR	InterPro; IPR003961; FN_III.
DR	InterPro; IPR008957; FN_III-like.
DR	InterPro; IPR005556; Glyco_hydro_48.
DR	InterPro; IPR008928; Glyco_trans_6hp.
DR	Pfam; PF00942; CBM_3; 1.
DR	Pfam; PF00041; fn3; 2.
DR	Pfam; PF02011; Glyco_hydro_48; 1.
DR	PRINTS; PR00844; GLHYDRLASE48.
DR	ProDom; PD001947; CBD_3; 1.
DR	ProDom; PD011903; Glyco_hydro_48; 2.
DR	SMART; SM00060; FN3; 2.
DR	PROSITE; PS0853; FN3; 2.
KW	Glycosidase; Hydrolase; Signal.
FT	SIGNAL 1 35
FT	CHAIN 36 1091
FT	SEQUENCE 1091 AA; 118000 MW; 21EACCEB2E704478 CRC64;
QY	Query Match 44.6%; Score 209.5; DB 2; Length 1091;
QY	Best Local Similarity 46.0%; Pred. No. 4.2e-14;
QY	Matches 40; Conservative 21; Mismatches 25; Indels 1; Gaps 1;
QY	2 SGGVKYQKNDSAPGDGNOIKGLQLVNTGSSVDLSTVTYRWFPRDGGSLVYNCOW 61
QY	940 TGTLEVYQYRNGSSAGSNAITQFNKNTGTTAIDLSKVKRYFTKD-SAADMSFWCDY 998
QY	62 AAMGCGNIRASFGSVNPATPTADTYLQ 88
QY	999 AOLGSANVQGSFVAVNPAGKTADTYLE 1025
QY	SEQUENCE 1091 AA; 118000 MW; 21EACCEB2E704478 CRC64;
QY	Query Match 44.6%; Score 209.5; DB 2; Length 1091;
QY	Best Local Similarity 46.0%; Pred. No. 4.2e-14;
QY	Matches 40; Conservative 21; Mismatches 25; Indels 1; Gaps 1;
QY	2 SGGVKYQKNDSAPGDGNOIKGLQLVNTGSSVDLSTVTYRWFPRDGGSLVYNCOW 61
QY	940 TGTLEVYQYRNGSSAGSNAITQFNKNTGTTAIDLSKVKRYFTKD-SAADMSFWCDY 998
QY	62 AAMGCGNIRASFGSVNPATPTADTYLQ 88
QY	999 AOLGSANVQGSFVAVNPAGKTADTYLE 1025
QY	SEQUENCE 1091 AA; 118000 MW; 21EACCEB2E704478 CRC64;
QY	Query Match 44.6%; Score 209.5; DB 2; Length 1091;
QY	Best Local Similarity 46.0%; Pred. No. 4.2e-14;
QY	Matches 40; Conservative 21; Mismatches 25; Indels 1; Gaps 1;
QY	2 SGGVKYQKNDSAPGDGNOIKGLQLVNTGSSVDLSTVTYRWFPRDGGSLVYNCOW 61
QY	940 TGTLEVYQYRNGSSAGSNAITQFNKNTGTTAIDLSKVKRYFTKD-SAADMSFWCDY 998
QY	62 AAMGCGNIRASFGSVNPATPTADTYLQ 88
QY	999 AOLGSANVQGSFVAVNPAGKTADTYLE 1025
QY	SEQUENCE 1091 AA; 118000 MW; 21EACCEB2E704478 CRC64;
QY	Query Match 44.6%; Score 209.5; DB 2; Length 1091;
QY	Best Local Similarity 46.0%; Pred. No. 4.2e-14;
QY	Matches 40; Conservative 21; Mismatches 25; Indels 1; Gaps 1;
QY	2 SGGVKYQKNDSAPGDGNOIKGLQLVNTGSSVDLSTVTYRWFPRDGGSLVYNCOW 61
QY	940 TGTLEVYQYRNGSSAGSNAITQFNKNTGTTAIDLSKVKRYFTKD-SAADMSFWCDY 998
QY	62 AAMGCGNIRASFGSVNPATPTADTYLQ 88
QY	999 AOLGSANVQGSFVAVNPAGKTADTYLE 1025
QY	SEQUENCE 1091 AA; 118000 MW; 21EACCEB2E704478 CRC64;
QY	Query Match 44.6%; Score 209.5; DB 2; Length 1091;
QY	Best Local Similarity 46.0%; Pred. No. 4.2e-14;
QY	Matches 40; Conservative 21; Mismatches 25; Indels 1; Gaps 1;
QY	2 SGGVKYQKNDSAPGDGNOIKGLQLVNTGSSVDLSTVTYRWFPRDGGSLVYNCOW 61
QY	940 TGTLEVYQYRNGSSAGSNAITQFNKNTGTTAIDLSKVKRYFTKD-SAADMSFWCDY 998
QY	62 AAMGCGNIRASFGSVNPATPTADTYLQ 88
QY	999 AOLGSANVQGSFVAVNPAGKTADTYLE 1025
QY	SEQUENCE 1091 AA; 118000 MW; 21EACCEB2E704478 CRC64;
QY	Query Match 44.6%; Score 209.5; DB 2; Length 1091;
QY	Best Local Similarity 46.0%; Pred. No. 4.2e-14;
QY	Matches 40; Conservative 21; Mismatches 25; Indels 1; Gaps 1;
QY	2 SGGVKYQKNDSAPGDGNOIKGLQLVNTGSSVDLSTVTYRWFPRDGGSLVYNCOW 61
QY	940 TGTLEVYQYRNGSSAGSNAITQFNKNTGTTAIDLSKVKRYFTKD-SAADMSFWCDY 998
QY	62 AAMGCGNIRASFGSVNPATPTADTYLQ 88
QY	999 AOLGSANVQGSFVAVNPAGKTADTYLE 1025
QY	SEQUENCE 1091 AA; 118000 MW; 21EACCEB2E704478 CRC64;
QY	Query Match 44.6%; Score 209.5; DB 2; Length 1091;
QY	Best Local Similarity 46.0%; Pred. No. 4.2e-14;
QY	Matches 40; Conservative 21; Mismatches 25; Indels 1; Gaps 1;
QY	2 SGGVKYQKNDSAPGDGNOIKGLQLVNTGSSVDLSTVTYRWFPRDGGSLVYNCOW 61
QY	940 TGTLEVYQYRNGSSAGSNAITQFNKNTGTTAIDLSKVKRYFTKD-SAADMSFWCDY 998
QY	62 AAMGCGNIRASFGSVNPATPTADTYLQ 88
QY	999 AOLGSANVQGSFVAVNPAGKTADTYLE 1025
QY	SEQUENCE 1091 AA; 118000 MW; 21EACCEB2E704478 CRC64;
QY	Query Match 44.6%; Score 209.5; DB 2; Length 1091;
QY	Best Local Similarity 46.0%; Pred. No. 4.2e-14;
QY	Matches 40; Conservative 21; Mismatches 25; Indels 1; Gaps 1;
QY	2 SGGVKYQKNDSAPGDGNOIKGLQLVNTGSSVDLSTVTYRWFPRDGGSLVYNCOW 61
QY	940 TGTLEVYQYRNGSSAGSNAITQFNKNTGTTAIDLSKVKRYFTKD-SAADMSFWCDY 998
QY	62 AAMGCGNIRASFGSVNPATPTADTYLQ 88
QY	999 AOLGSANVQGSFVAVNPAGKTADTYLE 1025
QY	SEQUENCE 1091 AA; 118000 MW; 21EACCEB2E704478 CRC64;
QY	Query Match 44.6%; Score 209.5; DB 2; Length 1091;
QY	Best Local Similarity 46.0%; Pred. No. 4.2e-14;
QY</	



```

DR InterPro; IPR008957; FN III-like.
DR InterPro; IPR001701; Glyco_hydro_9.
DR InterPro; IPR008928; Glyco_trans_6hp.
DR InterPro; IPR002345; Lipocalin.
DR Pfam; PF00942; CBM_3; 2.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF00759; Glyco_hydro_9; 1.
DR ProDom; PD001947; CBD_3; 2.
DR SMART; SM00060; FN3; 2.
DR PROSITE; PS00853; FN3; 2.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
KW Glycosidase; Hydrolase; Signal.
FT SIGNAL 1 34 Potential.
SQ SEQUENCE 997 AA; 106927 MW; F20CB119D6410629 CRC64;

Query Match 43.5%; Score 204.5; DB 2; Length 997;
Best Local Similarity 43.7%; Pred. No. 1.3e-13;
Matches 38; Conservative 21; Mismatches 27; Indels 1; Gaps 1;

QY 2 SGVKVQYKNDSPAGDNQIKPGLQVNTGSSVDLSLTVTVRYWFTRDGGSSTLVYNCW 61
DB 846 TGILEVYRSGSGSSNNAVTPQFNKNTGTQALDLSVKIRYFTKD-GTEELSFWDY 904

QY 62 AAMCGNIRASFGSVNPATPTADTYLQ 88
DB 905 AQGSANVQGMFVAVNPAKGTADTYVE 931

RESULT 8
QY 09AQG4 PRELIMINARY; PRT; 1751 AA.
AC 09AQG4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE CelE
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
RT thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
DR EMBL; AF078042; AAK06394.1; -.
DR HSSP; P37700; 1G87.
DR GO; GO:0004553; P:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul. bind.
DR InterPro; IPR001701; Glyco_hydro_9.
DR InterPro; IPR008928; Glyco_trans_6hp.
DR Pfam; PF00942; CBM_3; 4.
DR Pfam; PF00759; Glyco_hydro_9; 1.
DR ProDom; PD001947; CBD_3; 4.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
SQ SEQUENCE 1751 AA; 192176 MW; 60178CBF3C00BE95 CRC64;

Query Match 41.4%; Score 194.5; DB 2; Length 1751;
Best Local Similarity 44.7%; Pred. No. 3.3e-12;
Matches 38; Conservative 14; Mismatches 32; Indels 1; Gaps 1;

QY 4 GVKVQYKNDSPAGDNQIKPGLQVNTGSSVDLSLTVTVRYWFTRDGGSSTLVYNCW 63
DB 905 AQGSANVQGMFVAVNPAKGTADTYVE 931

Db 678 GVKVLYKNNETSASTGSRPMFKIVNGSSSSVDLSRVKIRYWTVDGDKPQSAV-CDWAQ 736
QY 64 MCGNIRASFGSVNPATPTADTYLQ 88
DB 737 IGASNVTFNFVKLSGSGVSGADYLYE 761

RESULT 9
QY 024820 PRELIMINARY; PRT; 1000 AA.
AC 024820;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Beta-glucanase.
OS thermophilic anaerobe NA10.
OC Bacteria.
OX NCBI_TaxID=67756;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NA10;
RA Miyake K., Machida Y., Hattori K., Iijima S.;
RL Submitted (Oct-1997) to the EMBL/Genbank/DBJ databases.
CC -!- SIMILARITY: Belongs to cellulase family F (family 10 of glycosyl
CC hydrolases).
DR EMBL; AB008029; BAA22939.1; -.
DR HSSP; Q06851; 1NEC.
DR GO; GO:0004553; P:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001589; Actbind_actnin.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul. bind.
DR InterPro; IPR001000; Glyco_hydro_10.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00942; CBM_3; 1.
DR Pfam; PF00150; Cellulase; 1.
DR Pfam; PF00331; Glyco_hydro_10; 1.
DR PRINTS; PR00134; GLHYDRLASE10.
DR ProDom; PD001947; CBD_3; 1.
DR SMART; SM00633; Glyco_10; 1.
DR PROSITE; PS00019; ACTININ_1; UNKNOWN_1.
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 1000 AA; 113264 MW; B9F659A56A752C6B CRC64;

Query Match 41.0%; Score 192.5; DB 2; Length 1000;
Best Local Similarity 43.7%; Pred. No. 2.9e-12;
Matches 38; Conservative 17; Mismatches 31; Indels 1; Gaps 1;

QY 2 SGVKVQYKNDSPAGDNQIKPGLQVNTGSSVDLSLTVTVRYWFTRDGGSSTLVYNCW 61
DB 372 SQGIKVLANKETNSTNTIRPWLKVNTGSSSIDLSRVITRYWTVYVDGDKAQSAYS-DW 430

QY 62 AAMCGNIRASFGSVNPATPTADTYLQ 88
DB 431 AQIGASNVTFNFVKLSGSGVSGADYLYE 457

RESULT 10
QY 09X3P5 PRELIMINARY; PRT; 1770 AA.
AC 09X3P5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE XynA.
GN Name-xynA;
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]

```







```
Db 413 GLKLVYKNNETSASTGSRPFWKIVNGSSVDLSRVKIRYWTYVDGKPKQSAV-CDWAQ 471
Qy 64 MCGNIRASFGSVNPTPTADTYLQ 88
Db 472 IGASNVTFNFVKLSSVSGADYILE 496

RESULT 13
P96311 PRELIMINARY; PRT; 1711 AA.
AC P96311;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE 1,4-beta-glucanase (Fragment).
GN Name=CeIA;
OS Anaerocellum thermophilum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Anaerocellum group;
OC Anaerocellum.
OX NCBI_TaxID=31899;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z-1320;
RX Zverlov V., Mahr S., Riedel K., Bronnenmeier K.;
RT "Properties and gene structure of a bifunctional cellulolytic enzyme (CeIA) from the extreme thermophile Anaerocellum thermophilum with separate glycosyl hydrolase family 9 and 48 catalytic domains.";
RL Microbiology 144:457-465(1998).
DR EMBL; Z86105; CAB06786.1; -.
DR PIR; T31337; T31337.
DR HSSP; P37700; 1687.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR000556; Glyco_hydro_48.
DR InterPro; IPR001701; Glyco_hydro_9.
DR InterPro; IPR008928; Glyco_trans_6hp.
DR InterPro; IPR002345; Lipocalin.
DR InterPro; IPR000209; Pept_S8_S53.
DR Pfam; PF00942; CBM_3; 3.
DR Pfam; PF02011; Glyco_hydro_48; 1.
DR Pfam; PF00753; Glyco_hydro_9; 1.
DR PRINTS; PR00844; GLHYDLASE48.
DR ProDom; PD001947; CBD_3; 3.
DR ProDom; PD011903; Glyco_hydro_48; 1.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 1711 AA; 169979 MW; E3E987CEB9CD0C21 CRC64;

Query Match 40.6%; Score 191; DB 2; Length 1711;
Best Local Similarity 43.8%; Pred. NO. 7.8e-12;
Matches 39; Conservative 18; Mismatches 30; Indels 2; Gaps 2;

Qy 1 VSGG-VKVOYKNDSPAGDNOIKPGLQLVNTGSSVDLSVTYVYFTRDGGSSTLVYNC 59
Db 683 VAGQIKVLYANKENSTNTIRPWLKIVNTGSSSIDLSRVIRIYWTYVDGKQAQSAIS- 741

Qy 60 DWAAAMCGNIRASFGSVNPTPTADTYLQ 88
Db 742 DWAGIGASNVTFKFKVLLSSVSGADYILE 770

RESULT 14
Q9AQHO PRELIMINARY; PRT; 996 AA.
AC Q9AQHO;
AC Q9AQHO;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
```

```
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Glycosyl hydrolase 5 (Fragment).
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
RT thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
DR EMBL; AF078038; AAK06388.1; -.
DR HSSP; Q06851; INEC.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR002860; Glyco_hydro_BNR.
DR InterPro; IPR011040; Sialidase.
DR Pfam; PF02012; BNR; 9.
DR Pfam; PF00942; CBM_3; 1.
DR ProDom; PD001947; CBD_3; 1.
DR Hydrolase. 996 AA; 108275 MW; 3C72B6ED22F3C614 CRC64;
KW NON_TER
FT NON_TER
SQ SEQUENCE 996 AA; 108275 MW; 3C72B6ED22F3C614 CRC64;

Query Match 40.5%; Score 190.5; DB 2; Length 996;
Best Local Similarity 43.5%; Pred. No. 4.7e-12;
Matches 37; Conservative 15; Mismatches 32; Indels 1; Gaps 1;

Qy 4 GVKVOYKNDSPAGDNOIKPGLQLVNTGSSVDLSVTYVYFTRDGGSSTLVYNCDDWA 63
Db 844 GLKLVYKNNETSASAGSRPFWKIVNGSSVDLSRVKIRYWTYVDGKPKQSAV-CDWAQ 902

Qy 64 MCGNIRASFGSVNPTPTADTYLQ 88
Db 903 IGASNVTFNFVKLSSVSGADYILE 927

RESULT 15
OS2374 PRELIMINARY; PRT; 1779 AA.
ID OS2374;
AC OS2374;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Family 10 xylanase (EC 3.2.1.8).
GN Name=xynC;
OS Caldicellulosiruptor sp. Rt69B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=70295;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rt69B.1;
RX MEDLINE=99283888; PubMed=10356996;
RA Morris D.D., Gibbs M.D., Ford M., Thomas J., Bergquist P.L.;
RA "Family 10 and 11 xylanase genes from Caldicellulosiruptor sp. strain
RT Rt69B.1.";
RL Extremophiles 3:103-111(1999).
CC -! SIMILARITY: Belongs to cellulase family F (family 10 of glycosyl
CC hydrolases).
DR EMBL; AF036924; AAB95326.1; -.
DR PIR; T31085; T31085.
DR HSSP; Q06851; INBC.
DR GO; GO:0030246; F:carbohydrate binding; IEA.
DR GO; GO:0004553; P:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR006584; CBD_IV.
```



```
DR InterPro; IPR005084; CBM_6.  
DR InterPro; IPR003305; CBM_CenC.  
DR InterPro; IPR008965; Cellul_bind.  
DR InterPro; IPR008979; Gal_bind_like.  
DR InterPro; IPR001000; Glyco_hydro_10.  
DR InterPro; IPR006710; Glyco_hydro_43.  
DR Pfam; PF00942; CBM_3; 3.  
DR Pfam; PF02018; CBM_4_9; 2.  
DR Pfam; PF03422; CBM_6; 1.  
DR Pfam; PF00331; Glyco_hydro_10; 1.  
DR Pfam; PF04616; Glyco_hydro_43; 1.  
DR PRINTS; PR00134; GLHYDRLASE10.  
DR PRODOM; PD001947; CBD_3; 3.  
DR SMART; SMO0606; CBD_IV; 1.  
DR SMART; SMO0633; Glyco_10; 1.  
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.  
KW Glicosidase, Hydrolase; Xylan degradation.  
SQ SEQUENCE      1779 AA;   194304 MW;    CE5269B606B5CED CRC64;  
  
Query Match          40.5%; Score 190.5; DB 2; Length 1779;  
Best Local Similarity 43.5%; Pred. No. 9.3e-12;  
Matches     37; Conservative 14; Mismatches 33; Indels 1; Gaps 1;  
  
Qy       4  GVRYOYKNNDSPGDNQIKPGLVNTGSSSVLSTVTWVRTRGGSSLTVNCDWA 63  
         :|::||::: |:::||::|::||::||::|:  
Db      1113 GLKVLKYNNETSASTGSIRPFVKIVNGSSVDLSRKVRKIYWTVTDDGDKPSAV-CDWAQ 1171
```

Search completed: October 5, 2005, 07:56:08  
Job time : 12.0656 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 07:49:28 ; Search time 3.41942 Seconds  
(without alignments)  
1942.949 Million cell updates/sec

Title: US-09-917-376-4

Perfect score: 470

Sequence: 1 VSGGVKQVQKNDSPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /cgn2\_6/ptodata/1/iaa/5A-COMB.pep:\*
- 2: /cgn2\_6/ptodata/1/iaa/5B-COMB.pep:\*
- 3: /cgn2\_6/ptodata/1/iaa/6A-COMB.pep:\*
- 4: /cgn2\_6/ptodata/1/iaa/6B-COMB.pep:\*
- 5: /cgn2\_6/ptodata/1/iaa/PCTUS-COMB.pep:\*
- 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	195.5	41.6	616	3	US-09-136-574A-47
2	194.5	41.4	1751	3	US-09-136-574A-44
3	191.5	40.7	1426	3	US-09-136-574A-43
4	172	36.6	700	2	US-07-862-588B-2
5	165.5	35.2	551	2	US-09-033-537A-1
6	165	35.1	167	5	PCT-US95-13813-9
7	165	35.1	476	4	US-09-339-159B-4
8	165	35.1	493	3	US-09-198-956-10
9	165	35.1	493	3	US-09-198-955A-12
10	165	35.1	493	3	US-09-694-531-12
11	165	35.1	493	3	US-09-670-141-10
12	165	35.1	493	4	US-10-072-152-12
13	155.5	33.1	1352	4	US-09-784-554B-2
14	150.5	32.0	1350	4	US-09-784-554B-4
15	120.5	25.6	531	2	US-07-862-588B-7
16	114	24.3	162	1	US-08-048-164A-2
17	114	24.3	162	1	US-08-460-462-2
18	114	24.3	162	1	US-08-460-457-2
19	114	24.3	162	1	US-08-460-458-2
20	114	24.3	162	2	US-08-460-455-2
21	114	24.3	162	2	US-08-330-394A-2
22	114	24.3	163	3	US-09-006-636-7
23	114	24.3	163	3	US-09-006-632-7
24	114	24.3	163	3	US-09-325-274-7
25	113	24.0	382	3	US-09-277-716-22
26	113	24.0	382	4	US-09-609-161B-22
27	112	23.8	154	2	US-08-330-394A-29

28	112	23.8	156	2	US-08-330-394A-22	Sequence 22, Appl
29	72.5	15.4	272	4	US-09-902-540-11056	Sequence 11056, A
30	65.5	13.9	1566	4	US-09-581-472B-2	Sequence 2, Appli
31	65	13.8	474	4	US-09-248-796A-25524	Sequence 25524, A
32	64	13.6	143	4	US-09-301-593-26	Sequence 26, Appl
33	64	13.6	428	3	US-09-118-319-5	Sequence 5, Appli
34	64	13.6	464	1	US-08-353-400-36	Sequence 36, Appl
35	64	13.6	472	4	US-09-301-593-30	Sequence 30, Appl
36	63.5	13.5	1785	3	US-09-341-587-3	Sequence 18, Appl
37	63	13.4	453	4	US-09-301-593-18	Sequence 43, Appl
38	63	13.4	472	4	US-09-301-593-43	Sequence 4640, Ap
39	63	13.4	718	4	US-09-328-352-4640	Sequence 12243, A
40	63	13.4	1133	4	US-09-902-540-12243	Sequence 2, Appli
41	63	13.4	1581	3	US-09-110-517-2	Sequence 38, Appl
42	62.5	13.3	288	3	US-09-423-439-38	Sequence 14347, A
43	62.5	13.3	389	4	US-09-902-540-14347	Sequence 33, Appl
44	62.5	13.3	445	1	US-08-353-400-33	Sequence 8, Appli
45	62.5	13.3	641	4	US-09-687-538B-8	

ALIGNMENTS

RESULT 1

US-09-136-574A-47

; Sequence 47, Application US/09136574A

; Patent No. 6294366

; GENERAL INFORMATION:

APPLICANT: Farrington, Graham K.

Anderson, Paige

Gibbs, Moreland

Bergquist, Peter

Daniels, Roy

Morgan, Hugh W.

Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for

Treating Cellulose Containing Fabrics Using Truncated

Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49

CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson

STREET: Spring House Corporate Center, P.O. Box 457

CITY: Spring House

STATE: PA

COUNTRY: USA

ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/136,574A

FILING DATE: 19-Aug-1998

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/932,571

FILING DATE: September 19, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Bak, Mary E.

REGISTRATION NUMBER: 31,215

REFERENCE/DOCKET NUMBER: 1997US001/CIP

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-540-9200

TELEFAX: 215-540-5818

TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 47:

SEQUENCE CHARACTERISTICS:

LENGTH: 616 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 47:



US-09-136-574A-47

Query Match 41.6%; Score 195.5; DB 3; Length 616;  
 Best Local Similarity 43.2%; Pred. No. 3.5e-13; Indels 1; Gaps 1;  
 Matches 38; Conservative 15; Mismatches 34;  
 QY 1 VSGGVKQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGKPKQSAV-CD 59  
 Db 1 MGSQVLYKNNETSASTGSRPWFKIVNGSSVDLSRVKIRYWTVDGKPKQSAV-CD 59  
 QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
 Db 60 WAQIGASNVTFNFVKLSSGSGADYLYE 87

RESULT 2

US-09-136-574A-44  
 ; Sequence 44, Application US/09136574A  
 ; Patent No. 6294366  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Farrington, Graham K.  
 ; Anderson, Paige  
 ; Gibbs, Moreland  
 ; Bergquist, Peter  
 ; Daniels, Roy  
 ; Morgan, Hugh W.  
 ; Williams, Diane P.  
 ; TITLE OF INVENTION: Compositions and Methods for  
 ; Treating Cellulose Containing Fabrics Using Truncated  
 ; Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Howson and Howson  
 STREET: Spring House Corporate Center, P.O. Box 457  
 CITY: Spring House  
 STATE: PA  
 COUNTRY: USA  
 ZIP: 19477

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/136,574A  
 FILING DATE: 19-Aug-1998  
 CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/932,571  
 FILING DATE: September 19, 1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Bak, Mary E.  
 REGISTRATION NUMBER: 31,215  
 REFERENCE/DOCKET NUMBER: 1997US001/CIP

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 215-540-9200  
 TELEFAX: 215-540-5818  
 TELEX: <Unknown>  
 INFORMATION FOR SEQ ID NO: 44:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1751 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear

MOLECULE TYPE: protein  
 SEQUENCE DESCRIPTION: SEQ ID NO: 44:  
 US-09-136-574A-44

Query Match 41.4%; Score 194.5; DB 3; Length 1751;  
 Best Local Similarity 44.7%; Pred. No. 1.6e-12;  
 Matches 38; Conservative 14; Mismatches 32; Indels 1; Gaps 1;

QY 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGKPKQSAV-CD 63

Db 678 GVKVLYKNNETSASTGSRPWFKIVNGSSVDLSRVKIRYWTVDGKPKQSAV-CD 736

QY 64 MCGNIRASFGSVNPATPTADTYLQ 88  
 Db 737 IGASNVTFNFVKLSSGSGADYLYE 761

RESULT 3

US-09-136-574A-43  
 ; Sequence 43, Application US/09136574A  
 ; Patent No. 6294366  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Farrington, Graham K.  
 ; Anderson, Paige  
 ; Gibbs, Moreland  
 ; Bergquist, Peter  
 ; Daniels, Roy  
 ; Morgan, Hugh W.  
 ; Williams, Diane P.  
 ; TITLE OF INVENTION: Compositions and Methods for  
 ; Treating Cellulose Containing Fabrics Using Truncated  
 ; Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Howson and Howson  
 STREET: Spring House Corporate Center, P.O. Box 457  
 CITY: Spring House  
 STATE: PA  
 COUNTRY: USA  
 ZIP: 19477

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/136,574A  
 FILING DATE: 19-Aug-1998  
 CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/932,571  
 FILING DATE: September 19, 1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Bak, Mary E.  
 REGISTRATION NUMBER: 31,215  
 REFERENCE/DOCKET NUMBER: 1997US001/CIP

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 215-540-9200  
 TELEFAX: 215-540-5818  
 TELEX: <Unknown>  
 INFORMATION FOR SEQ ID NO: 43:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1426 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear

MOLECULE TYPE: No. 6294366e  
 SEQUENCE DESCRIPTION: SEQ ID NO: 43:  
 US-09-136-574A-43

Query Match 40.7%; Score 191.5; DB 3; Length 1426;  
 Best Local Similarity 43.5%; Pred. No. 2.7e-12;  
 Matches 37; Conservative 15; Mismatches 32; Indels 1; Gaps 1;

QY 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGKPKQSAV-CD 63

Db 413 GLKVLKNNETSASTGSRPWFKIVNGSSVDLSRVKIRYWTVDGKPKQSAV-CD 471

QY 64 MCGNIRASFGSVNPATPTADTYLQ 88

Db 472 IGASNVTFNFVKLSSGSGADYLYE 496

RESULT 4



US-07-862-588B-2  
; Sequence 2, Application US/07862588B  
; Patent No. 5916796  
; GENERAL INFORMATION:  
; APPLICANT: Joergensen, Per Linna  
; APPLICANT: Sch lein, Martin  
; APPLICANT: Hansen, Christian  
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.  
; STREET: 405 Lexington Avenue, 62nd floor  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10017  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/862,588B  
; FILING DATE: 19920727  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DK 164/90  
; FILING DATE: 19-JAN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/DK91/00013  
; FILING DATE: 18-JAN-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Zelson, Steve T. / Lambiris, Elias J.  
; REGISTRATION NUMBER: 30,335 / 33,728  
; REFERENCE/DOCKET NUMBER: 3425.204-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212 867 0123  
; TELEFAX: 212 867 0298  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 700 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-07-862-588B-2  
Query Match 36.6%; Score 172; DB 2; Length 700;  
Best Local Similarity 44.9%; Pred. No. 1.6e-10;  
Matches 40; Conservative 18; Mismatches 27; Indels 4; Gaps 3;  
QY 1 VSGGVKQVKNND-SAPGDNQIKPGLQVNTGSSVDLSTVTVYWFTRDGGSSTLVYNC 59  
Db 549 VNSDLVQYKQDRNATDQIKPHFNQKGTSPVLSLTLRYFTKD-SSAAMNGWI 607  
QY 60 DWAMGCGNIRASFGSNPATPTADTYLQ 88  
Db 608 DWAKLGGSNQISFGNNGA--DSDTYAE 634  
RESULT 5  
US-09-033-537A-1  
; Sequence 1, Application US/09033537A  
; Patent No. 5958083  
; GENERAL INFORMATION:  
; APPLICANT: Onishi, Masahiro  
; APPLICANT: Fich, Merete  
; APPLICANT: Toft, Annette Hanne  
; APPLICANT: Sh lein, Martin  
; TITLE OF INVENTION: Prevention Of Back-Staining  
; TITLE OF INVENTION: In Stone Washing  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 5958083o No. 5958083disk of No. 5958083th America, Inc.  
; STREET: 405 Lexington Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: U.S.A.  
; ZIP: 10174  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/033,537A  
; FILING DATE: 02-MAR-1998  
; CLASSIFICATION: 008  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 0993/95  
; FILING DATE: 08-SEP-1995  
; APPLICATION NUMBER: PCT/DK96/00364  
; FILING DATE: 03-SEP-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Green, Reza  
; REGISTRATION NUMBER: 38,475  
; REFERENCE/DOCKET NUMBER: 4492.204-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-867-0123  
; TELEFAX: 212-878-9655  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 551 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-09-033-537A-1  
Query Match 35.2%; Score 165.5; DB 2; Length 551;  
Best Local Similarity 37.9%; Pred. No. 6e-10;  
Matches 33; Conservative 22; Mismatches 29; Indels 3; Gaps 2;  
QY 2 SGGVKVQYKNND-SAPGDNQIKPGLQVNTGSSVDLSTVTVYWFTRDGGSSTLVYNC 61  
Db 402 TGNLVVQYKVGDTSATDNQMKPFSFNKNNGTTPVNLGKLRYYFTKD-GTADMSASFDW 460  
QY 62 RAMGCGNIRASFGSNPATPTADTYLQ 88  
Db 461 AQIGASNVSAF--ANFTGSNTDTIVE 485  
RESULT 6  
PCT-US95-13813-9  
; Sequence 9, Application PC/TUS9513813  
; GENERAL INFORMATION:  
; APPLICANT: Yeda Research and Development Co. Ltd.  
; APPLICANT: Ramot University Authority for Applied  
; APPLICANT: Research and Industrial Development Ltd.  
; APPLICANT: Technion Research and Development Foundation Ltd.  
; APPLICANT: Bayer, Edward A.  
; APPLICANT: Morag, Ely  
; APPLICANT: Wilchek, Meir  
; APPLICANT: Lamed, Raphael  
; APPLICANT: Shoham, Yuval  
; TITLE OF INVENTION: MODIFIED CELLULOSE-BINDING DOMAIN (CBD)  
; TITLE OF INVENTION: PROTEINS AND USE THEREOF  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Browdy and Neimark  
; STREET: 419 Seventh Street N.W., Ste. 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM: disk  
; MEDIUM TYPE: Floppy disk







RESULT 11



```

; SEQ ID NO 12
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Clostridium thermocellum
US-10-072-152-12

Query Match      35.1%; Score 165; DB 4; Length 493;
Best Local Similarity 36.1%; Pred. No. 6e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGVKVQYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTVRYWFTRDGSSSLVYVNC 60
Db 331 VSGNLKVEFYNSPDDTTNSINPQKVTNTGSSAIDLSKLTLYRYVTVDGQKQTFW-CD 389

QY 61 WAAM-----GCGNIRASFGSVNPATPTADTYLQ 88
Db 390 HAALIGNSYNGITSNVKGFVRKMSSTNNADTYLE 426

RESULT 13
US-09-784-554B-2
; Sequence 2, Application US/09784554B
; Patent No. 6815192
; GENERAL INFORMATION:
; APPLICANT: Schnorr, Kirk
; APPLICANT: Jorgensen, Per Lina
; APPLICANT: Schulein, Martin
; TITLE OF INVENTION: FAMILY 44 XYLOGLUCANASES
; FILE REFERENCE: 10017.200-US
; CURRENT APPLICATION NUMBER: US/09/784,554B
; CURRENT FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1352
; TYPE: PRT
; ORGANISM: Paenibacillus polymyxa
US-09-784-554B-2

Query Match      33.1%; Score 155.5; DB 4; Length 1352;
Best Local Similarity 31.8%; Pred. No. 2.3e-08;
Matches 27; Conservative 21; Mismatches 36; Indels 1; Gaps 1;

QY 4 GVKVQYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTVRYWFTRDGSSSLVYVNC 63
Db 1200 GLLQYRTADTKVNDNHLNPHQFQVKNKGTSPINELKIRYYITDGDREQ-TFNC 1258

QY 64 MCGNIRASFGSVNPATPTADTYLQ 88
Db 1259 LSCSKLNGKLVKMKAKATGADYYLE 1283

RESULT 14
US-09-784-554B-4
; Sequence 4, Application US/09784554B
; Patent No. 6815192
; GENERAL INFORMATION:
; APPLICANT: Schnorr, Kirk
; APPLICANT: Jorgensen, Per Lina
; APPLICANT: Schulein, Martin
; TITLE OF INVENTION: FAMILY 44 XYLOGLUCANASES
; FILE REFERENCE: 10017.200-US
; CURRENT APPLICATION NUMBER: US/09/784,554B
; CURRENT FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 1350
; TYPE: PRT
; ORGANISM: Paenibacillus polymyxa
US-09-784-554B-4

Query Match      32.0%; Score 150.5; DB 4; Length 1350;

```

```

Best Local Similarity 31.8%; Pred. No. 8.1e-08;
Matches 27; Conservative 21; Mismatches 36; Indels 1; Gaps 1;

QY 4 GVKVQYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTVRYWFTRDGSSSLVYVNC 63
Db 1199 GLVQYRTADTKVNDNHLNPHQFQVKNKGTSPINELKIRYYITDGDREQ-TFNC 1257

QY 64 MCGNIRASFGSVNPATPTADTYLQ 88
Db 1258 LSCSKLNGKLVKMKAKATGADYYLE 1282

RESULT 15
US-07-862-588B-7
; Sequence 7, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linaa
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 59167960 No. 5916796disk of No. 5916796ch America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 531 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-862-588B-7

Query Match      25.6%; Score 120.5; DB 2; Length 531;
Best Local Similarity 32.9%; Pred. No. 5e-05;
Matches 28; Conservative 18; Mismatches 36; Indels 3; Gaps 2;

QY 2 SGGVQYKYNDSAPGDNQIKPGLQVNTGSSVDLSTVTVRYWFTRDGSSSLVYVNC 61
Db 432 TGNLVVQYKVGDTSATDNQKPSFNKKNKGTTPVNLGKLVKLVKLVKLVKLVKLV 490

QY 62 AMGCGNIRASFGSVNPATPTADTY 86
Db 491 AQIGRTNVLAF--ANFTGSNTDITY 513

Search completed: October 5, 2005, 08:04:31

```



Job time : 4.41942 secs

---



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 07:52:23 ; Search time 11.3031 Seconds  
(without alignments)  
3268.602 Million cell updates/sec

Title: US-09-917-376-4  
Perfect score: 470  
Sequence: 1 VSGGVKQVYKNNDSAPGDNQ.....RASFGVNPATPTADTYLQX 89

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1846076 seqs, 415116000 residues

Total number of hits satisfying chosen parameters: 1846076

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA.\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US10E\_PUBCOMB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/1/pubpaa/US11A\_PUBCOMB.pep.\*  
20: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*  
21: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
22: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	469	99.8	88	11	US-09-917-376-5
2	469	99.8	88	14	US-10-155-400-5
3	469	99.8	89	11	US-09-917-376-4
4	469	99.8	89	14	US-10-155-400-4
5	469	99.8	154	10	US-09-917-378-4
6	469	99.8	762	10	US-09-917-378-1
7	466	99.1	150	10	US-09-917-384-5
8	466	99.1	150	10	US-09-917-383-5
9	466	99.1	1043	10	US-09-917-384-6
10	466	99.1	1043	10	US-09-917-383-6
11	466	99.1	1228	10	US-09-917-384-1

12	466	99.1	1228	10	US-09-917-383-1	Sequence 1, Appli
13	462	98.3	957	11	US-09-917-376-1	Sequence 1, Appli
14	462	98.3	957	14	US-10-155-400-1	Sequence 1, Appli
15	257	54.7	741	14	US-10-156-761-8100	Sequence 8100, Ap
16	190	40.4	508	15	US-10-369-493-23151	Sequence 23151, A
17	188	40.0	477	16	US-10-466-208-12	Sequence 12, Appli
18	188	40.0	496	16	US-10-466-208-8	Sequence 8, Appli
19	188	40.0	677	16	US-10-433-577-35	Sequence 35, Appli
20	180	38.3	1621	14	US-10-185-990-10	Sequence 10, Appli
21	165	35.1	167	17	US-10-933-404-4	Sequence 4, Appli
22	165	35.1	476	15	US-10-372-054-4	Sequence 4, Appli
23	165	35.1	493	13	US-10-072-152-12	Sequence 12, Appli
24	165	35.1	493	15	US-10-655-433-12	Sequence 12, Appli
25	165	35.1	599	10	US-09-955-555A-29	Sequence 29, Appli
26	155.5	33.1	1352	10	US-09-784-554B-2	Sequence 2, Appli
27	155.5	33.1	1352	16	US-10-896-555-2	Sequence 2, Appli
28	150.5	32.0	1350	10	US-09-784-554B-4	Sequence 4, Appli
29	150.5	32.0	1350	16	US-10-896-555-4	Sequence 4, Appli
30	130.5	27.8	1483	15	US-10-282-122A-51483	Sequence 51483, A
31	114	24.3	256	14	US-10-261-446-6	Sequence 6, Appli
32	114	24.3	256	15	US-10-261-445B-6	Sequence 6, Appli
33	114	24.3	256	17	US-10-782-234-6	Sequence 6, Appli
34	113	24.0	382	10	US-09-808-898-22	Sequence 22, Appli
35	105.5	22.4	163	15	US-10-460-524-2	Sequence 2, Appli
36	78.5	16.7	1049	15	US-10-282-122A-49900	Sequence 49900, A
37	75.5	16.1	346	16	US-10-437-963-179234	Sequence 223, App
38	75.5	16.1	618	15	US-10-211-462-223	Sequence 223, App
39	75.5	16.1	618	16	US-10-723-860-4039	Sequence 4039, Ap
40	75.5	16.1	618	18	US-10-756-149-5713	Sequence 5713, Ap
41	75.5	16.1	986	9	US-09-747-835A-53	Sequence 53, Appli
42	75.5	16.1	986	14	US-10-120-604-101	Sequence 101, App
43	75.5	16.1	986	14	US-10-225-567A-406	Sequence 406, App
44	75.5	16.1	986	15	US-10-398-458-3	Sequence 3, Appli
45	75.5	16.1	986	15	US-10-312-312-53	Sequence 53, Appli

ALIGNMENTS

RESULT 1  
US-09-917-376-5  
; Sequence 5, Application US/09917376  
; Publication No. US20040038334A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: 40197.4US01  
; CURRENT APPLICATION NUMBER: US/09/917,376  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 88  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; OTHER INFORMATION: Carbohydrate binding domain  
US-09-917-376-5

Query Match	99.8%	Score 469;	DB 11;	Length 88;
Best Local Similarity	100.0%;	Pred. No. 3.6e-48;		
Matches	88;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
OY	1	VSGGVKQVYKNNDSAPGDNQIKFGLQLVNTGSSVDLSTVTVYVWFTDGGSSFLVNC	60	
Db	1	VSGGVKQVYKNNDSAPGDNQIKFGLQLVNTGSSVDLSTVTVYVWFTDGGSSFLVNC	60	
OY	61	WAAMGCGNIRASFGSVNPATPTADTYLQ	88	



```

Db 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
RESULT 2
US-10-155-400-5
; Sequence 5, Application US/10155400
; Publication No. US20030108988A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: NREL 01-36A
; CURRENT APPLICATION NUMBER: US/10/155,400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 88
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; OTHER INFORMATION: Carbohydrate binding domain
US-10-155-400-5
Query Match 99.8%; Score 469; DB 14; Length 88;
Best Local Similarity 100.0%; Pred. No. 3.6e-48;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVKYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTVRYWFTTRDGGSSTLVYNCD 60
Db 1 VSGGVKQVKYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTVRYWFTTRDGGSSTLVYNCD 60

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
Db 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
RESULT 3
US-09-917-376-4
; Sequence 4, Application US/09917376
; Publication No. US20040038334A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40197.4US01
; CURRENT APPLICATION NUMBER: US/09/917,376
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 89
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: {89}
; OTHER INFORMATION: Any amino acid
; OTHER INFORMATION: Carbohydrate binding domain
US-09-917-376-4
Query Match 99.8%; Score 469; DB 11; Length 89;
Best Local Similarity 100.0%; Pred. No. 3.6e-48;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVKYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTVRYWFTTRDGGSSTLVYNCD 60
Db 1 VSGGVKQVKYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTVRYWFTTRDGGSSTLVYNCD 60

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
Db 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
RESULT 4
US-09-917-378-4
; Sequence 4, Application US/09917378
; Publication No. US20030119093A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT MANNANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40197.7US01
; CURRENT APPLICATION NUMBER: US/09/917,378
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Carbohydrate
; OTHER INFORMATION: binding domain
US-09-917-378-4
Query Match 99.8%; Score 469; DB 14; Length 89;
Best Local Similarity 100.0%; Pred. No. 3.6e-48;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVKYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTVRYWFTTRDGGSSTLVYNCD 60
Db 1 VSGGVKQVKYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTVRYWFTTRDGGSSTLVYNCD 60

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
Db 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
RESULT 5
US-09-917-378-4
; Sequence 4, Application US/09917378
; Publication No. US20030119093A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT MANNANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40197.7US01
; CURRENT APPLICATION NUMBER: US/09/917,378
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Carbohydrate
; OTHER INFORMATION: binding domain
US-09-917-378-4
Query Match 99.8%; Score 469; DB 14; Length 89;
Best Local Similarity 100.0%; Pred. No. 3.6e-48;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVKYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTVRYWFTTRDGGSSTLVYNCD 60
Db 1 VSGGVKQVKYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTVRYWFTTRDGGSSTLVYNCD 60

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
Db 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88

```



Query Match 99.8%; Score 469; DB 10; Length 154;  
 Best Local Similarity 100.0%; Pred. No. 7e-48;  
 Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VSGGVKVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTTRDGGSSSTLYVNC 60  
 Db 1 VSGGVKVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTTRDGGSSSTLYVNC 60  
 QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
 Db 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88

RESULT 6  
 US-09-917-378-1  
 ; Sequence 1, Application US/09917378  
 ; Publication No. US20030119093A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DING, SHI-YOU  
 ; APPLICANT: ADNEY, WILLIAM S.  
 ; APPLICANT: VINZANT, TODD B.  
 ; APPLICANT: DECKER, STEPHEN R.  
 ; APPLICANT: HIMMEL, MICHAEL E.  
 ; TITLE OF INVENTION: THERMAL TOLERANT MANNANASE FROM ACIDOTHERMUS  
 ; TITLE OF INVENTION: CELLULOXYTICUS  
 ; FILE REFERENCE: 40197.7US01  
 ; CURRENT APPLICATION NUMBER: US/09/917,378  
 ; CURRENT FILING DATE: 2001-07-28  
 ; NUMBER OF SEQ ID NOS: 8  
 ; SOFTWARE: Patent In Ver. 2.1  
 ; SEQ ID NO 1  
 ; TYPE: PRT  
 ; LENGTH: 762  
 ; ORGANISM: Acidothermus cellulolyticus  
 US-09-917-378-1

Query Match 99.8%; Score 469; DB 10; Length 762;  
 Best Local Similarity 100.0%; Pred. No. 4.9e-47;  
 Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VSGGVKVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTTRDGGSSSTLYVNC 60  
 Db 455 VSGGVKVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTTRDGGSSSTLYVNC 514  
 QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
 Db 515 WAAMCGNIRASFGSVNPATPTADTYLQ 542

RESULT 7  
 US-09-917-384-5  
 ; Sequence 5, Application US/09917384  
 ; Publication No. US20030096342A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DING, SHI-YOU  
 ; APPLICANT: ADNEY, WILLIAM S.  
 ; APPLICANT: VINZANT, TODD B.  
 ; APPLICANT: DECKER, STEPHEN R.  
 ; APPLICANT: HIMMEL, MICHAEL E.  
 ; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
 ; TITLE OF INVENTION: CELLULOXYTICUS  
 ; FILE REFERENCE: 40170.6US01  
 ; CURRENT APPLICATION NUMBER: US/09/917,384  
 ; CURRENT FILING DATE: 2001-07-28  
 ; NUMBER OF SEQ ID NOS: 14  
 ; SOFTWARE: Patent In Ver. 2.1  
 ; SEQ ID NO 5  
 ; LENGTH: 150  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: Segment of

OTHER INFORMATION: Guxa  
 US-09-917-384-5  
 Query Match 99.1%; Score 466; DB 10; Length 150;  
 Best Local Similarity 98.9%; Pred. No. 1.6e-47;  
 Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VSGGVKVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTTRDGGSSSTLYVNC 60  
 Db 1 VSGGLKVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTTRDGGSSSTLYVNC 60  
 QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
 Db 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88

RESULT 8  
 US-09-917-383-5  
 ; Sequence 5, Application US/09917383  
 ; Publication No. US20030104522A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DING, SHI-YOU  
 ; APPLICANT: ADNEY, WILLIAM S.  
 ; APPLICANT: VINZANT, TODD B.  
 ; APPLICANT: DECKER, STEPHEN R.  
 ; APPLICANT: HIMMEL, MICHAEL E.  
 ; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
 ; TITLE OF INVENTION: CELLULOXYTICUS  
 ; FILE REFERENCE: 40170.6US01  
 ; CURRENT APPLICATION NUMBER: US/09/917,383  
 ; CURRENT FILING DATE: 2001-07-28  
 ; NUMBER OF SEQ ID NOS: 14  
 ; SOFTWARE: Patent In Ver. 2.1  
 ; SEQ ID NO 5  
 ; LENGTH: 150  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: Segment of  
 ; OTHER INFORMATION: Guxa  
 US-09-917-383-5

Query Match 99.1%; Score 466; DB 10; Length 150;  
 Best Local Similarity 98.9%; Pred. No. 1.6e-47;  
 Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VSGGVKVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTTRDGGSSSTLYVNC 60  
 Db 1 VSGGLKVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTTRDGGSSSTLYVNC 60  
 QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
 Db 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88

RESULT 9  
 US-09-917-384-6  
 ; Sequence 6, Application US/09917384  
 ; Publication No. US20030096342A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DING, SHI-YOU  
 ; APPLICANT: ADNEY, WILLIAM S.  
 ; APPLICANT: VINZANT, TODD B.  
 ; APPLICANT: DECKER, STEPHEN R.  
 ; APPLICANT: HIMMEL, MICHAEL E.  
 ; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
 ; TITLE OF INVENTION: CELLULOXYTICUS  
 ; FILE REFERENCE: 40170.6US01  
 ; CURRENT APPLICATION NUMBER: US/09/917,384  
 ; CURRENT FILING DATE: 2001-07-28  
 ; NUMBER OF SEQ ID NOS: 14  
 ; SOFTWARE: Patent In Ver. 2.1  
 ; SEQ ID NO 6



```
; LENGTH: 1043
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-384-6

Query Match          99.1%; Score 466; DB 10; Length 1043;
Best Local Similarity 98.9%; Pred. No. 1.6e-46;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLYVNC 60
Db 477 VSGGLKVQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLYVNC 536

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
Db 537 WAAMCGNIRASFGSVNPATPTADTYLQ 564

RESULT 10
US-09-917-383-6
; Sequence 6, Application US/09917383
; Publication No. US20030104522A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,383
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1043
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-383-6

Query Match          99.1%; Score 466; DB 10; Length 1043;
Best Local Similarity 98.9%; Pred. No. 1.6e-46;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLYVNC 60
Db 477 VSGGLKVQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLYVNC 536

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
Db 537 WAAMCGNIRASFGSVNPATPTADTYLQ 564

RESULT 11
US-09-917-384-1
; Sequence 1, Application US/09917384
; Publication No. US20030096342A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40170.6US01
```

```
; CURRENT APPLICATION NUMBER: US/09/917,384
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1228
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-384-1

Query Match          99.1%; Score 466; DB 10; Length 1228;
Best Local Similarity 98.9%; Pred. No. 2e-46;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLYVNC 60
Db 584 VSGGLKVQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLYVNC 643

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
Db 644 WAAMCGNIRASFGSVNPATPTADTYLQ 671

RESULT 12
US-09-917-383-1
; Sequence 1, Application US/09917383
; Publication No. US20030104522A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,383
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1228
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-383-1

Query Match          99.1%; Score 466; DB 10; Length 1228;
Best Local Similarity 98.9%; Pred. No. 2e-46;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLYVNC 60
Db 584 VSGGLKVQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLYVNC 643

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
Db 644 WAAMCGNIRASFGSVNPATPTADTYLQ 671

RESULT 13
US-09-917-376-1
; Sequence 1, Application US/09917376
; Publication No. US20040038344A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
```



Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-273697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 8100  
LENGTH: 741  
TYPE: PRT  
ORGANISM: Streptomyces avermitilis  
US-10-156-761-8100  
Query Match 54.7%; Score 257; DB 14; Length 741;  
Best Local Similarity 56.3%; Pred. No. 1.2e-21;  
Matches 49; Conservative 13; Mismatches 25; Indels 0; Gaps 0;  
QY 2 SGGVKVQYKNDSSAPGDNQIKPGLQVNTGSSVDLSTVTYVYFTRDGGSSSTLYVNCW 61  
DB 591 SGGKLVLYKNDSSATDNAIKPGIRIVNTGSSVDLSKVTARYYFTRDGGSPYVNAWCDY 650  
QY 62 AAMCGNIRASFGSVNPATPTADTYLQ 88  
DB 651 AAVGCSNVSLKVVPLTTPVPGADAYLE 677  
Search completed: October 5, 2005, 08:08:32  
Job time : 12.3031 secs

APPLICANT: HIMMEL, MICHAEL E.  
TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
TITLE OF INVENTION: CELLULOYTICUS  
FILE REFERENCE: 40197.4US01  
CURRENT APPLICATION NUMBER: US/09/917,376  
CURRENT FILING DATE: 2001-07-28  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 957  
TYPE: PRT  
ORGANISM: Acidothermus cellulolyticus  
FEATURE:  
NAME/KEY: MOD RES  
LOCATION: (957)  
OTHER INFORMATION: Any amino acid  
US-09-917-376-1  
Query Match 98.3%; Score 462; DB 11; Length 957;  
Best Local Similarity 97.7%; Pred. No. 4.4e-46;  
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VSGGVKQYKNDSSAPGDNQIKPGLQVNTGSSVDLSTVTYVYFTRDGGSSSTLYVNC 60  
DB 869 VSGGVKQYKNDSSAPGDNQIKPGLQVNTGSSVDLSTVTYVYFTRDGGSSSTLYVNC 928  
QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
DB 929 WAAIGCGNIRASFGSVNPATPTADTYLQ 956  
RESULT 14  
US-10-155-400-1  
Sequence 1, Application US/10155400  
Publication No. US20030108988A1  
GENERAL INFORMATION:  
APPLICANT: DING, SHI-YOU  
APPLICANT: ADNEY, WILLIAM S.  
APPLICANT: VINZANT, TODD B.  
APPLICANT: HIMMEL, MICHAEL E.  
TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
TITLE OF INVENTION: CELLULOYTICUS  
FILE REFERENCE: NREL 01-36A  
CURRENT APPLICATION NUMBER: US/10/155,400  
CURRENT FILING DATE: 2002-10-22  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 957  
TYPE: PRT  
ORGANISM: Acidothermus cellulolyticus  
FEATURE:  
NAME/KEY: MOD RES  
LOCATION: (957)  
OTHER INFORMATION: Any amino acid  
US-10-155-400-1  
Query Match 98.3%; Score 462; DB 14; Length 957;  
Best Local Similarity 97.7%; Pred. No. 4.4e-46;  
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VSGGVKQYKNDSSAPGDNQIKPGLQVNTGSSVDLSTVTYVYFTRDGGSSSTLYVNC 60  
DB 869 VSGGVKQYKNDSSAPGDNQIKPGLQVNTGSSVDLSTVTYVYFTRDGGSSSTLYVNC 928  
QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
DB 929 WAAIGCGNIRASFGSVNPATPTADTYLQ 956  
RESULT 15  
US-10-156-761-8100  
Sequence 8100, Application US/10156761







```
Query Match: 99.79% Indels: 0
DB: 6 Gaps: 0
US-09-917-376-4 (1-89) x AX700050 (1-2289)

Qy 1 ValSerGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 1363 GTGTCCGGTGGGTGAGGTGCAGTACAAAGCAATGATTCGGCGCGGGTGATAACCCAG 1422

Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1423 ATCAACCGGCTCCAGTTGGTGAATACGGGGTGTGCTCGGTGGATTTCACCGGTG 1482

Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerThrLeuValTyrAsnCysAsp 60
Db 1483 ACGTGGCGTACTGGTTCACCGGGATGGTGGTCTGTCGACACTGGTGACACTGTGAC 1542

Qy 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 1543 TGGCGCGGATGGGTGTGGGAATATCCGCGCTCGTTCGGCTCGGTGAACCCGCGACG 1602

Qy 81 ProThrAlaAspThrTyrLeuGln 88
Db 1603 CCGACGGCGACACCTACCTGCAG 1626

RESULT 2
AX700036
LOCUS AX700036 3687 bp DNA linear PAT 03-APR-2003
DEFINITION Sequence 2 from Patent WO03012109.
ACCESSION AX700036
VERSION AX700036.1 GI:29536019
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
LOCATION/Qualifiers
1. :3687
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Segment of Guxa"

ORIGIN
Alignment Scores:
Pred. No.: 1.55e-47 Length: 3687
Score: 466.00 Matches: 87
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 98.86% Mismatches: 0
Query Match: 99.15% Indels: 0
Gaps: 0

US-09-917-376-4 (1-89) x AX700036 (1-3687)

Qy 1 ValSerGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 1750 GTGTCCGGTGGGTGAGGTGCAGTACAAAGCAATGATTCGGCGCGGGTGATAACCCAG 1809

Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1810 ATCAACCGGCTCCAGTTGGTGAATACCGGGTGTGCTCGGTGGATTTCACCGGTG 1869

Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerThrLeuValTyrAsnCysAsp 60
Db 1870 ACGTGGCGTACTGGTTCACCGGGATGGTGGTCTGTCGACACTGGTGACACTGTGAC 1929

Qy 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 1930 TGGCGCGGATGGGTGTGGGAATATCCGCGCTCGTTCGGCTCGGTGAACCCGCGACG 1989

Qy 81 ProThrAlaAspThrTyrLeuGln 88
Db 1990 CCGACGGCGGACACCTACCTGCAG 2013

RESULT 3
AX700058
LOCUS AX700058 2869 bp DNA linear PAT 03-APR-2003
DEFINITION Sequence 2 from Patent WO03012090.
ACCESSION AX700058
VERSION AX700058.1 GI:29536021
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
LOCATION/Qualifiers
1. :2869
/organism="Acidothermus cellulolyticus"
/mol_type="unassigned DNA"
/db_xref="taxon:28049"

ORIGIN
Alignment Scores:
Pred. No.: 3.55e-47 Length: 2869
Score: 462.00 Matches: 86
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 97.73% Mismatches: 0
Query Match: 98.30% Indels: 0
Gaps: 0

US-09-917-376-4 (1-89) x AX700058 (1-2869)

Qy 1 ValSerGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 2605 GTGTCCGGTGGGTGAGGTGCAGTACAAAGCAATGATTCGGCGCGGGTGATAATCAG 2664

Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 2665 ATCAACCGGCTTCGAGGTGTGATACCGGGTGTGCTCGGTGGATTTCGACGGTG 2724

Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerThrLeuValTyrAsnCysAsp 60
Db 2725 ACGGTGGGTACTGGTTTCAACCGGATGGTGGCTCGTGCACACTGGTGACACTGTGAC 2784

Qy 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 2785 TGGCGCGGATGGGTGTGGGAATATCCGCGCTCGTTCGGCTCGGTGAACCCGCGACG 2844

Qy 81 ProThrAlaAspThrTyrLeuGln 88
Db 2845 CCGACGGCGGACACCTACCTGCAG 2868

RESULT 4
AX700025
LOCUS AX700025 3365 bp DNA linear PAT 03-APR-2003
DEFINITION Sequence 2 from Patent WO03012095.
ACCESSION AX700025
VERSION AX700025.1 GI:29536018
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
```



JOURNAL	Patent: WO 03012095-A 2 13-FEB-2003; Midwest Research Institute (US)	TITLE	Sakaki, Y. and Hattori, M.
FEATURES	source	JOURNAL	Direct Submission
	1. .3365 /organism="Acidothermus cellulolyticus" /mol_type="unassigned DNA" /db_xref="taxon:28049"		Submitted (29-MAR-2002) Director-General of Biotechnology Center, National Institute of Technology and Evaluation, Biotechnology Center, 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan (E-mail:bio@nitech.go.jp, URL:http://www.bio.nitech.go.jp/, Tel:81-3-3481-1933, Fax:81-3-3481-8424)
ORIGIN		COMMENT	This work was done in collaboration with Haruo Ikeda (*1), Jun Ishikawa (*2), Akiharu Hanamoto (*3), Chigusa Takahashi (*3), Mayumi Shinose (*3), Hiroshi Horikawa (*4), Hidekazu Nakazawa (*4), Tomomi Osonoe (*4), Norihiro Kushida (*4), Hisashi Kikuchi (*4), Tadayoshi Shiba (*5), Yoshiyuki Sakaki (*6, *7), Masahira Hattori (*1, *7) and Satoshi Omura (*1, *3). Final finishing process and all annotation were done by H. Ikeda and J. Ishikawa.
Alignment Scores:			*1 Kitasato Institute for Life Sciences, Kitasato University *2 National Institute of Infectious Diseases *3 The Kitasato Institute *4 National Institute of Technology and Evaluation *5 School of Science, Kitasato University *6 Institute of Medical Science, University of Tokyo *7 RIKEN, Genomic Sciences Center
Pred. No.:	7.55e-44		Following url is also available. http://avermitilis.is.kitasato-u.ac.jp.
Score:	436.00		Location/Qualifiers
Percent Similarity:	97.62%		1. .299175
Best Local Similarity:	96.43%		/organism="Streptomyces avermitilis MA-4680"
Query Match:	92.77%		/mol_type="genomic DNA"
DB:	6		/strain="MA-4680"
			/db_xref="taxon:227882"
US-09-917-376-4 (1-89) x AX700025 (1-3365)			/note="This strain is also named as strain: ATCC 31267, NCIMB 12804 or NRRL 8165.-synonym: Streptomyces avermitilinus"
QY 5 VallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleIysProGly 24		FEATURES	complement(155. .550)
Db 112 CTCAAAGCGCAGTATAAGAACAAATGTCGGCGCGAGTGACACCGATCAACCGGCT 171		source	/note="SAV445"
QY 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyr 44			Tn4561, truncated"
Db 172 CTCAGTGGTGAATACCGCGGTCGTCGGTGGATTTGTCCGCGTACCGTGGCGGTAC 231			/codon start=1
QY 45 TrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAlaMet 64			/transl_table=11
Db 232 TGGTTCACCGGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 291			/product="putative transposase"
QY 65 GlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAlaAsp 84			/protein_id="BAC68155.1"
Db 292 GGGTGTGGCAATATCCGGCGCTCGTTCGGTCTCGTGGTGGTGGTGGTGGTGGTGGT 351			/db_xref="GI:29604084"
QY 85 ThrTyrLeuGln 88		CDS	/translation="MAGPPRSATRSRTTADREDSAPHPADRADEPGYRRQIKVOANLQ EGHALLARRAGOLYORQDMEDQICAGLGLNALVLFNTRYMDAALTQLRADGFEVR DEDVARLSPFMHSHINVLGRYSFQLPDLR"
Db 352 ACTACCTGCAG 363			complement(474. .740)
RESULT 5			/note="IS630 family, truncated SAV446"
AP005023	299175 bp DNA linear BCT 10-MAY-2003		/codon start=1
LOCUS	Streptomyces avermitilis genomic DNA, complete genome, section		/transl_table=11
DEFINITION	3/30.		/product="putative transposase"
ACCESSION	AP005023		/protein_id="BAC68156.1"
VERSION	AP005023.1		/db_xref="GI:29604085"
KEYWORDS	Streptomyces avermitilis MA-4680		/translation="MPDTONAGFSTVTAAAVSTAGSVSMAPTSPVMAPPSPGRRR EFGSLGAVRAYDVIRMLRSDGRPTPLGDAITYYGGSRRLCTSCA"
SOURCE	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.		766. .987
ORGANISM	Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M.		/note="SAV447"
REFERENCE	1		/codon start=1
AUTHORS	Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M.		/transl_table=11
TITLE	Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites		/product="putative transposase"
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)		/protein_id="BAC68157.1"
MEDLINE	21477403		/db_xref="GI:29604086"
PUBMED	11572948		/translation="WVVAVGCALTSVPTAGGVAVPVCVRLTDSQGLKQIVRRG SASSVRFRRAMMLASAGGNRPVIAQLV"
REFERENCE	2		complement(1954. .2571)
AUTHORS	Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M. and Omura, S.		/note="SAV448"
TITLE	Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis		/codon start=1
JOURNAL	Nat. Biotechnol. 21 (5), 526-531 (2003)		/transl_table=11
MEDLINE	22608306		/product="putative Tetr-family transcriptional regulator"
PUBMED	12692562		/protein_id="BAC68158.1"
REFERENCE	3 (bases 1 to 299175)		/db_xref="GI:29604087"
AUTHORS	Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kushida, N., Director-General of Biotechnology Center, Shiba, T.,		/translation="MPTSTRRARERANTRRIIEAALHVLTEGIAALTIRRIATDVE YSPVYVQHPANKDALVELVAHGRMLSEFQAQEPDTRMTRIASQYVRFAGE HPHLFQVMDPMDADERRRVAEPVIDVLKELTSSAAHDVVLADFDQACEILWGTL"



```
YGIASLYSGVGNDRARYLAEOALHTLILGWTETTPANGRPASS"
2694. .3098
/note="SAV449"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68159.1"
/db_xref="GI:29604088"
/translation="MIYHINRVTKATATPEQIEGVLESWRNQSRNPAIKSFVVGNN
HGSDYTVGAVFVVEHLDGLFAYLTHTPTTYOTDHLGLHVERLEIFDVSDDDDDPDLNKA
IQELHRRNELNPQIAGMLADVPTTGSVD"
3127. .4029
/note="SAV450"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68160.1"
/db_xref="GI:29604089"
/translation="MAKHVIIGAGSIGTNVARLLAEGESVRIIVTRSGSGPEHPLVDR
VAADASPSRLTELISRGAGVITYHCANPSYTWERLLPPLQTAATIAKANDAVIALT
GSLVAYGPOQGRMNEHTPMAATGHKGLRRRMWEOALAGIRTVVRGSDYIGKDN
GIYSLIKSLAKGKAWI TGHLDMPHTPTFNGDMAOALITLASEERAWGRPWHPVSA
PAVITRELARYAVAAGRPKLIQIPRSVTRTAGLIVPIARENAEMDYQWYAPFMD
ATETTFGTATDLDFAVRNEVG"
complement (4130. .4498)
/gene="bldG"
complement (4130. .4498)
/gene="bldG"
/note="SAV451"
/codon_start=1
/transl_table=11
/product="putative anti-anti-sigma factor"
/protein_id="BAC68161.1"
/db_xref="GI:29604090"
/translation="MTDAFHDVHTDAALAVIALSGEFDDIAAAPAVPARALELIANG
HPDLVADLSGVTFCDSSGLGALVIGWRCAKDAGSLTAAIPDLRSLLSVTGMDAFL
PAYSSADAALAARQNRRTA"
4635. .5156
/note="SAV452"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68162.1"
/db_xref="GI:29604091"
/translation="MFAVKSRVRVGEIATPTGEALELVLHGESLRGVRESSDA
TAPDGLDFLAELPARGEWHTALVRSPVEGEACGLFHPGAEPSGEPARRLCWRE
SAPRTIDIEPVEVHTLRRSLIEDLIGAPAHLRPAAPRTHGRRRCPLVHGPVRPLAAV
LVHGTAAGPCTPR"
5183. .5665
/note="SAV453"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68163.1"
/db_xref="GI:29604092"
/translation="MVTAVRARAGSAWQACRGARRSSOHSSEGVRRMDEKEFLRTV
AERTGLSQERADITWATLETLARLSPGEARDLITLPEGLAEAVRGTGTDRIERFG
HHDSVVRVAERTMLKEBADRGVRTVLAVLREALISEKFNDLMSQLGTDIFSQALESAG
"
5810. .6964
/note="SAV454"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68164.1"
/db_xref="GI:29604093"
/translation="MIWTADVPTSSGRSVRVVALVADPGLPSEIARALAPRLPHLR
AVGARVWQDVTWTPALSVSEQADVSDIAENVRSHLDGAEWDVAIFLTPRRRLYP
ISVEDTALRAALISLPALGVRRRLRRVRQAVDVVRELVAHESTPRHPVIGRPPE
FVPTDRPGPQOPQVAVVPGRLGHTRLVAGMYANRPWLFGSLSRALAGVPATVLF
INSATNWLALQVAVOHVITAGSAVALTLMWIVDHLMERGTGLPARHPLFYPLV
TLVTIALAVVLAALFATLVAVSFLLLDASVLKFI GROVGVGDYFLFAWFIATAIM
IGGAFGTSLGDEKVSADAAVYGRQRQRQRMRLQMPRPERGERGPDITDG"

complement (6986. .7498)
/note="SAV455"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68165.1"
/db_xref="GI:29604094"
/translation="MNPYAPASGIEQORFGWMLDEYLMCSDIIADLAERGRLEQRPLV
IADLARAGERVLAVGPAQRVTRALGDRYMHQNAVAFGSPFVLGSMASWATSALG
NSARRAAANDVAQPRCGARTPGLHVLVDLVHGTMLRWSMHLGTSSTRPSSALRAILL
ELVDLHVIAL"
7548. .8189
/note="SAV456"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68166.1"
/db_xref="GI:29604095"
/translation="MATLLRRVCPHEGVRIIAVPAAGFVLSRRPWTASVTRCAVVGPOS
EDREVESRTFVASASSRLRPSVRCPRRAVPSQAVRSRFPRTGTRFLGDLV
VYQRRDALHARVHRVIDVHPGLGRGAARPVRIATHSLGGVIAVDMATANVPLWTES
LVTFGSQATPFHQPVPGDLLCQMIAFRMRKHQEAETGALPPTGTFAHAGQGW"

Alignment Scores:
Pred. No.: 4,31e-19 Length: 299175
Score: 257.00 Matches: 49
Percent Similarity: 71.26% Conservative: 13
Best Local Similarity: 56.32% Mismatches: 25
Query Match: 54.68% Indels: 0
DB: 1 Gaps: 0

US-09-917-376-4 (1-89) x AP005023 (1-299175)

Qy 2 SerGlyGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspGlnIle 21
Db 105341 TCGGCGCGCCCTCAAGGTCCTCTACAAAGAACACAGACTCTCTCGGCCACCCGACACGCCATC 105400
Qy 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrValThr 41
Db 105401 CGGCCAGCGCTTCGATCGTCAACACCGCGACGGCTCCTCGACCTGTCCAAGGTCAAG 105460
Qy 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
Db 105461 GCCCGCTACTACTTCACCGCGACAGCGCTCGCCACCGTGAACCGCTGTCGCGACTAC 105520
Qy 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81
Db 105521 CGCGCGCTCGGCTGTTCACACGTCAGCTGAAGGTGTTACCCCTGACCCCTGACCCGCCGCGTGC 105580
Qy 82 ThrAlaAspThrTyrLeuGln 88
Db 105581 GGAGCCGACGCTACTCTGAA 105601

RESULT 6
AY298814
LOCUS Thermobifida fusca strain TM51 endoglucanase (cel5B) gene, complete
DEFINITION cds.
ACCESSION AY298814
VERSION AY298814.1 GI:31745732
KEYWORDS
SOURCE Thermobifida fusca
ORGANISM Thermobifida fusca
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptoporangineae; Nocardiopsaceae; Thermobifida.
REFERENCE
1 (bases 1 to 1957)
Posta,K., Beki,E., Kukolya,J. and Hornok,L.
Phylogenetic relationships of Tf cel5B, a new endoglucanase
encoding gene from Thermobifida fusca
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 1957)
Posta,K., Beki,E., Kukolya,J. and Hornok,L.
Direct Submission
TITLE
```



JOURNAL Submitted (14-MAY-2003) Agricultural Biotechnology and Microbiology, Szent Istvan University Godollo, Pater Karoly str.1., Godollo 2103, Hungary

FEATURES Location/Qualifiers  
1..1957  
/organism="Thermobifida fusca"  
/mol\_type="genomic DNA"  
/strain="TW51"  
/db\_xref="taxon:2021"

misc\_binding  
26..39  
/note="putative"  
/bound\_moiety="CelR"  
107..1957  
/gene="cel5B"  
107..1957  
/gene="cel5B"  
/codon\_start=1  
/transl\_table=11  
/product="endoglucanase"  
/protein\_id="AAP56348.1"  
/db\_xref="GI:31745733"

gene  
107..1957  
/gene="cel5B"  
107..1957  
/gene="cel5B"  
/codon\_start=1  
/transl\_table=11  
/product="endoglucanase"  
/protein\_id="AAP56348.1"  
/db\_xref="GI:31745733"  
/translation="MTPLRLRAGAAAIAGASALIPLTSPAAASGTADWLHTDGN  
RIVDSAGNEVLTGANWFENFTSERMFHGLWAANIEDITSMAERGINNVRVPISTQL  
LLEKNGQAGSVNVEYNPELAGMNTLEVDYWLQCEYGLKVMLDVHSAEADNSG  
HYYPWYKGDITTEDFYTAWEWTERYKNNDTIVAADIKNRPHKANETPRKWDGST  
DIDNFKVCETAGKRILALINPNMLILCGRIEYKDGQDWSGDRDYSTWVGNNLR  
GVADHPDGAHQDLVSPHDYSPVFEQFEGEWNKQITLEDVWRPNWLYIHEDD  
IAPLILGSGGFGDLGDNKQWMTALRSIIIDKMHHTFWALNPNSGDTGGLNLYDWT  
WDEAKYFLKALWQDANGKFGVLDHVDPLGSGVSTGVSINQYGGGPGSPPTPEPT  
EPPTPEPTPEPTPEPTPEPTPEPTPEPTPEPTPEPTPEPTPEPTPEPTPEPTPEPT  
VEIHYFYNEPGGTILQFTCDMAQVCANVNASFTSLAPGADTSLVLTLSGLAPGAS  
TELQRIHTANWANFDESDDYSRGNTDNLSEVITAYLGLTLVWGTPPA"  
236..1258  
/gene="cel5B"  
/note="Region: catalytic domain; GH5 family"

misc\_feature  
1532..1765  
/gene="cel5B"  
/note="Region: cellulose binding domain; CBM-3"

ORIGIN  
Alignment Scores:  
Pred. No.: 2,78e-18 Length: 1957  
Score: 228.50 Matches: 45  
Percent Similarity: 70.93% Conservative: 16  
Best Local Similarity: 52.33% Mismatches: 22  
Query Match: 48.62% Indels: 3  
DB: 1 Gaps: 2

US-09-917-376-4 (1-89) x AY298814 (1-1957)

QY 2 SerGlyGlyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21  
1517 ACGGTGGCTCGAGTCTTACCGCAACACAGCTTTCGGCGCCGACGACCCAGATC 1576

QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41  
1577 GCGCGGGGCTGGCTCGGTCTGCTCAACACCGAAGCAGCAGCTGCTGCGTGGGAA 1636

QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTtp 61  
1637 ATCCACTACTTCTACCAACAGCCGCGCGC---GGTACCCTTCAGTTCACCTCGACTGG 1693

QY 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81  
1694 GCTCAAGTGGGTGGCCCAACGCTCAACGGCTCTCTTCACGTCGCTG-----TCGGCTCCG 1747

QY 82 ThrAlaAspThrTyrLeu 87  
1748 GCGCGCGACACCTCCCTG 1765

RESULT 7  
AF163837  
LOCUS AF163837 4567 bp DNA linear BCT 08-FEB-2000

DEFINITION Caldicibacillus cellulovorans multidomain beta-1,4-mannanase precursor (mana) gene, complete cds; and unknown genes.  
ACCESSION AF163837  
VERSION AF163837.1 GI:6651325

KEYWORDS Caldicibacillus cellulovorans  
SOURCE Caldicibacillus cellulovorans  
ORGANISM Bacteria; Firmicutes; Bacillales; Alicyclobacillaceae; Caldicibacillus.

REFERENCE  
AUTHORS 1 (bases 1 to 4567)  
TITLE Sunna,A., Gibbs,M.D., Chin,C.W., Nelson,P.J. and Bergquist,P.L.  
A gene encoding a novel multidomain beta-1,4-mannanase from Caldicibacillus cellulovorans and action of the recombinant enzyme on kraft pulp

JOURNAL Appl. Environ. Microbiol. 66 (2), 664-670 (2000)  
MEDLINE 20120520  
PUBMED 10653733

REFERENCE 2 (bases 1 to 4567)  
AUTHORS Sunna,A., Gibbs,M.D. and Bergquist,P.L.  
TITLE Direct Submission

JOURNAL Submitted (29-JUN-1999) Biological Sciences, Macquarie University, North Ryde, New South Wales 2109, Australia  
FEATURES Location/Qualifiers  
1..4567  
/organism="Caldicibacillus cellulovorans"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:74586"

CDS  
1..513  
/note="orf1"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="GI:6651326"  
/db\_xref="GI:6651326"

misc\_feature  
1..60  
/note="Region: proline-threonine linker"  
/evidence=not\_experimental  
61..510  
/note="Region: cellulose-binding domain type IIib"

misc\_feature  
620..3412  
/evidence=not\_experimental  
620..3412  
/gene="mana"  
620..3412  
/gene="mana"  
/codon\_start=1  
/transl\_table=11  
/product="multidomain beta-1,4-mannanase precursor"

misc\_feature  
61..510  
/note="Region: cellulose-binding domain type IIib"

gene  
620..3412  
/evidence=not\_experimental  
620..3412  
/gene="mana"

CDS  
620..3412  
/evidence=not\_experimental  
620..3412  
/gene="mana"

misc\_feature  
1..60  
/note="Region: proline-threonine linker"

misc\_feature  
61..510  
/note="Region: cellulose-binding domain type IIib"

gene  
620..3412  
/evidence=not\_experimental  
620..3412  
/gene="mana"

CDS  
620..3412  
/evidence=not\_experimental  
620..3412  
/gene="mana"

misc\_feature  
620..3409  
/note="encodes multidomain beta-1,4-mannanase precursor"







misc_feature	/product="beta-1,4-xylanase XynA"		source
483..965			
misc_feature	/gene="xynA"		gene
/note="Region: thermostabilizing domain"			
966..1946			promoter
/gene="xynA"			
misc_feature	/notes="Region: beta-1,4-xylanase catalytic domain"		-10_signal
1947..2075			
misc_feature	/gene="xynA"		RBS
/note="Region: proline-threonine linker"			
2076..2519			CDS
/gene="xynA"			
misc_feature	/notes="Region: cellulose-binding domain typeIIb"		
2520..2696			
misc_feature	/notes="Region: proline-threonine linker"		
2697..3146			
misc_feature	/gene="xynA"		
/note="Region: cellulose-binding domain typeIIb"			
ORIGIN			
Alignment Scores:			
Pred. No.:	2.91e-16	Length: 3237	
Score:	214.50	Matches: 40	
Percent Similarity:	65.12%	Conservative: 16	
Best Local Similarity:	46.51%	Mismatches: 29	
Query Match:	45.64%	Indels: 1	
DB:	1	Gaps: 1	
US-09-917-376-4 (1-89) x AF200304 (1-3237)			
QY	3 GlyGlyVallysValGlnTyrIysAsnAsnAspSerAlaProGlyAspAsnGlnIleIys 22		
Ddb	2694 GGCACCTGGTCTGTCAGTACCGCGCGGACACGACGCGGCGACACCGCTGAAG 2753		
QY	23 ProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrVal 42		
Ddb	2754 CGCATTTTCGATGTTGAACCGCGGACGACGAGCGTACCGTTGTCGAGCTTACGATT 2813		
QY	43 ArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAla 62		
Ddb	2814 CGGTACTGGTACACGGTGGAC---GGCGACAGCCGACGAGGTGTTCACTGTGACTGGGCG 2870		
QY	63 AlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThr 82		
Ddb	2871 TGGTGGAGTTCGAACCTGCGGCGAGTCTGGTGAAGTTGACGCGGCGGCGGCGG 2930		
QY	83 AlaAspThrTyrLeuGln 88		
Ddb	2931 GCGGACTACTACCTTGAG 2948		
RESULT 9			
PAE488933	3509 bp DNA linear BCT 15-JUN-2002		
LOCUS	Paenibacillus sp. BP-23 cel48C gene for cellulose		
DEFINITION	1,4-beta-cellobiosidase.		
ACCESSION	AJ488933		
VERSION	AJ488933.1 GI:21449823		
KEYWORDS	cel48C gene; cellulose 1,4-beta-cellobiosidase.		
SOURCE	Paenibacillus sp. BP-23		
ORGANISM	Paenibacillus sp. BP-23		
REFERENCE	Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.		
AUTHORS	1		
TITLE	Sanchez,M.M., Pastor,F.I.J. and Diaz,P.		
JOURNAL	Paenibacillus sp. BP-23 family 48 cellulase. Cloning and		
REFERENCE	performance on cellulosic substrates		
AUTHORS	Unpublished		
TITLE	2 (bases 1 to 3509)		
JOURNAL	Diaz,P.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (31-MAY-2002) Diaz P., Microbiology, University of		
TITLE	Barcelona, Av. Diagonal, 645, 08028-Barcelona, SPAIN		
JOURNAL	Location/Qualifiers		
FEATURES			
1..3509			
/organism="Paenibacillus sp. BP-23"			
/mol_type="genomic DNA"			
/strain="BP-23"			
/db_xref="taxon:198119"			
/country="Spain:Catalonia, Ebro river delta"			
20..3472			
/gene="cel48C"			
20..25			
/gene="cel48C"			
20..25			
/gene="cel48C"			
39..45			
/gene="cel48C"			
62..3337			
/gene="cel48C"			
/EC number="3.2.1.91"			
/function="cellulose hydrolysis"			
/codon_start=1			
/transl_table=11			
/product="cellulose 1,4-beta-cellobiosidase"			
/protein_id="CAD32945.1"			
/db_xref="GI:21449824"			
/db_xref="GOA:Q8KKF7"			
/db_xref="UniProt/TREMBL:Q8KKF7"			
/translation="MLKTAAKSLTAMLAGITVMTLTGTLGTSQTAAHSDQAIEIOA			
DGINEARFLQLYAOLKDPANGYFSAEGIPYHSIETLLSEAPDYGHMSTSEAYSYMLWL			
ETMYGHTYGDMSKLEAANDMEKYIIPVNEGDNREOPTMSKYNPNSPATYAAKEKFP			
DYPNSNLGQYAAGKDPDLDAELKATYGNQTYLMEWLIDVDNNGYGYGLNLSHTSTY			
VNTFQGEQSVWEAIPHSQDDKSPKAGEGFMSLFTKESQAPSAQWRITNATDADA			
RAVQVLYWAKEMGNNBYLDRAKMGDFLRYGMDYKVPQVSKAKNGTPTPGTKDS			
NMYLMWYTSWGGGLGQGDWAWRIGASHTHQAYQNPVAAVALSDPAGGLIPDSPYAK			
ADMNATLKRQLEFYTWLOSHEGAVGSGATNSIGSSYAAYPAGVSTFYDMAYOEAQVVR			
DPDSNWFQFQANPLERVAEMYIILAESGDLTSENFOAKKVTITKWDWTKDVFEVGE			
RPVTDAGQYILNAAQGRILGGTNAQVATPAPGEPWIPGCGQEQWQGPDKWGFSTYTN			
NPNFHAITKDPQDVTGVLGSIKALTFFAAGTEAENGTLISAKQEAQDLAQLSLDTAM			
DYDNGVGIYTBEEERKDYFRYFAKEIYLPANWSGTFGQNTIETGAGVPSPDPAKGGNG			
YIGYSLDPAIKQDPMAVLDNKYKTSYNTPTTKWENGAPFTYHRSFWSQDMATAYG			
SFVATLGSSTFGQVQVPSAPAGVTAAGDKQVVLWNAAAGATYSVKRATVNGPVT			
SVATVGTSTFTDGLTNGTATYVITAVNSAGESAPSAQVATPAGVTPGVFTLS			
GMAGDAQAVLAWNTASTGAATYVQRSTGTGAYANLATGLTALTYITDTTAVNGTAYNR			
VVTAVASQTNNSNVVLTTPMAAPISTGTLEVYRNGSSAGNAITPQNLKNTGTTA			
IDLSSKVRYYFTKDSADMSFCVDAQLGSANVQGSFVAVNPAGKTADTLYLELSFA			
GAGSLAAGAEFTGIIQTRFSKNNWTNFDQSDNYSFNATQTAFGAWTKVTAQDGAKVWG			
MEP"			
62..166			
/gene="cel48C"			
167..3334			
/gene="cel48C"			
/product="cellulose 1,4-beta-cellobiosidase"			
/EC number="3.2.1.91"			
3456..3472			
/gene="cel48C"			
sig_peptide			
mat_peptide			
terminator			
ORIGIN			
Alignment Scores:			
Pred. No.:	1.35e-15	Length: 3509	
Score:	209.50	Matches: 40	
Percent Similarity:	70.11%	Conservative: 21	
Best Local Similarity:	45.98%	Mismatches: 25	
Query Match:	44.57%	Indels: 1	
DB:	1	Gaps: 1	
US-09-917-376-4 (1-89) x PAE488933 (1-3509)			
QY	2 SerGlyGlyVallysValGlnTyrIysAsnAsnAspSerAlaProGlyAspAsnGlnIle 21		
Ddb	2879 ACGGGAACCGTGGAGTGCAGTATCCGATTCGAGGTCCAGTTCAGTAATGCAATT 2938		
QY	22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41		
Ddb	2939 ACGCGCAATTAACTCTGAAATAATACAGGTACACGCAATTGATCTGAGTGAAGTGAAG 2998		



```

Qy 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTyr 61
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2999 GTTCGTTATTATTTACGAGGAC---AGCGCAGCGGACATGAGCTTCTGCTGTGACTAT 3055

Qy 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3056 GCACAGCTTGGCAGTGCATGTCAGCGGCGAGCTTGTGCTGTGTAAATCCGGGAGGCG 3115

Qy 82 ThrAlaAspThrTyrLeuGln 88
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3116 ACGCAGATACGTATCTGGAG 3136

RESULT 10
BSP133614
LOCUS BSP133614 Bacillus sp. BP23 celB gene. 4161 bp DNA linear BCT 09-MAY-2001
DEFINITION AUI33614
ACCESSION AUI33614
VERSION AUI33614.1 GI:4490765
KEYWORDS celB gene; cellulase; glycosyl hydrolase.
SOURCE Bacillus sp. BP-23
ORGANISM Bacillus sp. BP-23
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

REFERENCE
1
AUTHORS Pastor, F.I., Pujol, X., Blanco, A., Vidal, T., Torres, A.L. and Diaz, P.
TITLE Molecular cloning and characterization of a multidomain
endoglucanase from Paenibacillus sp BP-23: evaluation of its
performance in pulp refining
JOURNAL Appl. Microbiol. Biotechnol. 55 (1), 61-68 (2001)
MEDLINE 21129642
PUBMED 11234960
REFERENCE
2 (bases 1 to 4161)
AUTHORS Diaz, P.
DIRECT SUBMISSION
TITLE Submitted (10-MAR-1999) Diaz P., Microbiology, University of
Barcelona, Av.Diagonal 645, Barcelona-08028, SPAIN
JOURNAL Location/Qualifiers
FEATURES
source
1. 4161
/organism="Bacillus sp. BP-23"
/mol_type="genomic DNA"
/strain="BP-23"
/db_xref="taxon:89769"
398..3525
/gene="celB"
398..404
/gene="celB"
437..442
/gene="celB"
514..519
/gene="celB"
532..3525
/gene="celB"
/EC_number="3.2.1.4"
/standard_name="endoglucanase"
/codon_start=1
/transl_table=11
/product="cellulase"
/protein_id="CAB38941.1"
/db_xref="GI:4490766"
/db_xref="GOA:Q92411"
/db_xref="UniProt/TrEMBL:Q92411"
/translation="MKGSWRRVAIVTLASGLAGFSMTNLRQADAAAGDHNYAEA
LQKATYFEQRSGPLPANNRVEWRGDSMGQDGVDTLGGWYDAGDHVKFGPWA
ASATMLAWSVEVYADGHEQAGQLEEIKDNIKWATDYPKHAHTKRNELWQVAGNTH
AWMGPAEWMQNRSPFKIDACPGSDLAETAALAAASSIVFADSDPVSIAKLQJHAK
ELYNFADTYRGKYDCITDRAAFYNSWTGYEDELAWGAWLYLATNDSDPVLKALSAA
DRWSTSGSANWPYTWQWDSKHYGAQILLARITSLNMPKATKPIQSTERNLDYWT
VGTNGGRVKTYPGGLWLDOWGSLRYAANAAFISFYSDWVSDPDKSRVONFATSOI
NYILGDNRQSSVYVGGQNSPQHPHRTAHSWMNEDI PANHRHLLYKAWGCGPNA
SDQITDDIGDVSNEVATDYNAGFTGALKNNLLYQNHQHPANFPAPVPEVDFYVE
AAVRSNGSYTEIRALLNRRSGWPMRGDQLSFYFLDLSEVYAAGRTVSDVQVTVSS
SEGATVSPVVDAAKRIYAITANFSNTKIYPGGEGNYRKEVQPIRTGPGQAMNPAND
PSYQNLITGNPKSNIIPVVDAGVKVSGQBPVTPVAVPAAPAGVQAVAGNSQVALNW
SASAGASVTVKRAEVSQGPPTTAAAGVNGLTYTNTGLTNGKTYTYVTVAVNSAGESP

```

```

ASVQVSGMPQAATTVPGAUTLSTAGTANNQNNMLSWTAASGASSYTVQRAVAGTYYTDA
TGLAVLNYNDTTLNKGTSYSYRVAANASGQTLNNVMLTSPGSPVSTGLEVRSG
GSGNSNAVTPOFLNKGTTQAIIDLSTVKIRYFTKGTGTEBLSFDCDVAQVGSANVQ
MFVAVNPAKGTADTVVEISFTSGAGSLAAGAETGVIQTRFSKNNWSAFDQSNDSYDA
SKTAPAAWKKVTAYQGNTOVWGLEP"
532..633
/gene="celB"

ORIGIN
Alignment Scores:
Pred. No.: 7.03e-15 Length: 4161
Score: 204.50 Matches: 38
Percent Similarity: 67.82% Conservative: 21
Best Local Similarity: 43.68% Mismatches: 27
Query Match: 43.51% Indels: 1
DB: 1 Gaps: 1

US-09-917-376-4 (1-89) x BSP133614 (1-4161)
Qy 2 SerGlyGlyValIysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGlnIle 21
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3067 ACAGTACGCTTGAGTGCAGTATCGAAGCGGGGTTTCAGGCAATTCAGCAATGCAGTT 3126

Qy 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3127 ACTCGCAATTCATCTGAAGATACCGGCACACAGCGGATTCGATCTGAGTACGGTGAAA 3186

Qy 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTyr 61
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3187 ATTCGTTATTACTTTACCAAGAC---GGCACGGAGGAGCTGCTCTTCGTGTGATTAT 3243

Qy 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3244 GCGCAGTGGGCTCAGCGAAGCTACAGGAATGTTGTAGCAGTGAATCCGCGGAAGGT 3303

Qy 82 ThrAlaAspThrTyrLeuGln 88
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3304 ACAGCGGATACGTATGTGGAG 3324

RESULT 11
E35142 2029 bp DNA linear PAT 18-JUN-2001
LOCUS E35142
DEFINITION Truncated cellulase composition.
ACCESSION E35142
VERSION E35142.1 GI:13018967
KEYWORDS JP 1999221086-A/44.
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1 (bases 1 to 2029)
AUTHORS Paiji, A., Petaer, L.B., Roy, M.D., Gurahamu, K.F., Moreland, D.G.,
Hyu, M. and Daian, P.W.
Truncated cellulase composition
Patent: JP 1999221086-A 44 17-AUG-1999;
CLARIANT INTERNATIONAL LTD
OS Artificial Sequence
PN JP 1999221086-A/44
PF 17-AUG-1999
PD 19-SEP-1998 JP 1998283606
PR 19-SEP-1997 US 08/932571
PI PAIJI ANDERSON, PETAER L BAGUKUISUTO, ROY M DANIEL, PI
GURAHAMU K PARINTON,
PI MORELAND DAVID GIBUSU, HYU MORGAN, DAIAN PURATONIORISU WILLIAM
PC C12N15/09, C12D3/386, C12N1/21, C12N9/42// (C12N1/21, C12R1:19), PC
(C12N9/42, C12R1:19), C12N15/00
CC
FH Key Location/Qualifiers
FT source 1..2029 /organism='Artificial Sequence'.
FEATURES
source 1..2029
/organism="synthetic construct"
/mol_type="genomic DNA"

```



/db\_xref="taxon:32630"

## ORIGIN

Alignment Scores:  
Pred. No.: 3,78e-14 Length: 2029  
Score: 195.50 Matches: 38  
Percent Similarity: 60.23% Conservativeness: 15  
Best Local Similarity: 43.18% Mismatches: 34  
Query Match: 41.60% Indels: 1  
DB: 6 Gaps: 1

US-09-917-376-4 (1-89) x E35142 (1-2029)

QY 1 ValSerGlyGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
DB 1 ATGGGAAGTGGTGAAGTACTGTACAAAGCAATGAGACAAAGTGCAGCAGGTTCT 60  
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
DB 61 ATAGCGCGTGGTTTAAAGATAGTGAATGGAGGCAGCAGGTTGATCTTAGCAGGTT 120  
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
DB 121 AAGTAAGACTACTGTACACAGTGGTGTGACAAAGCCACAGAGTGGCGTA---TGTGAC 177  
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
DB 178 TGGGCACAGATAGGGGCAAGCAATGTGACATTCAATTTTGTGAAGCTTAGCAGCGGAGTG 237  
QY 81 ProThrAlaAspThrTyrLeuGln 88  
DB 238 AGTGGAGCGGATTATTACTCTGGAG 261

## RESULT 12

E35143 LOCUS 2029 bp DNA linear PAT 18-JUN-2001  
DEFINITION Truncated cellulase composition.  
ACCESSION E35143  
VERSION E35143.1 GI:13018968  
KEYWORDS JP 1999221086-A/45.  
SOURCE unclassified  
ORGANISM unclassified  
REFERENCE 1 (bases 1 to 2029)  
AUTHORS Paiji, A., Petaer, L.B., Roy, M.D., Gurahamu, K.F., Moreland, D.G.,  
Hyu, M. and Dajan, P.W.  
TITLE Truncated cellulase composition  
JOURNAL Patent: JP 199221086-A 45 17-AUG-1999;  
COMMENT CLARIANT INTERNATIONAL LTD  
OS Unidentified  
PN JP 1999221086-A/45  
PD 17-AUG-1999  
PF 21-SEP-1998 JP 1998283606  
PR 19-SEP-1997 US 08/952571  
PI PAIJI ANDERSON, PETAER L BAGUKUISUTO, ROY M DANIEL, PI  
GURAHAMU K FARINTON,  
PI MORELAND DAVID GIBUSU, HYU MORGAN, DAJAN PURATONIORISU WILLIAM  
PC C12N15/09, C12N15/386, C12N1/21, C12N9/42, C12N1/21, C12R1/19, PC  
(C12N9/42, C12R1/19), C12N15/00  
CC

Key Location/Qualifiers  
FT source 1. 2029  
FT /organism='Unidentified'.

## FEATURES

source  
1. 2029 Location/Qualifiers  
/organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"

## ORIGIN

Alignment Scores:  
Pred. No.: 3,78e-14 Length: 2029  
Score: 195.50 Matches: 38

Percent Similarity: 60.23% Conservativeness: 15  
Best Local Similarity: 43.18% Mismatches: 34  
Query Match: 41.60% Indels: 1  
DB: 6 Gaps: 1

US-09-917-376-4 (1-89) x E35143 (1-2029)

QY 1 ValSerGlyGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
DB 1 ATGGGAAGTGGTGAAGTACTGTACAAAGCAATGAGACAAAGTGCAGCAGGTTCT 60  
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
DB 61 ATAGCGCGTGGTTTAAAGATAGTGAATGGAGGCAGCAGGTTGATCTTAGCAGGTT 120  
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
DB 121 AAGTAAGACTACTGTGACACAGTGGTGTGACAAAGCCACAGAGTGGCGTA---TGTGAC 177  
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
DB 178 TGGGCACAGATAGGGGCAAGCAATGTGACATTCAATTTTGTGAAGCTTAGCAGCGGAGTG 237  
QY 81 ProThrAlaAspThrTyrLeuGln 88  
DB 238 AGTGGAGCGGATTATTACTCTGGAG 261

## RESULT 13

AF0780385 LOCUS 6005 bp DNA linear BCT 11-FEB-2001  
DEFINITION Caldicellulosiruptor sp. Tok7B.1 glycosyl hydrolase 6 gene, partial  
cgs: and Cels gene, complete cds.  
ACCESSION AF078042  
VERSION AF078042.1 GI:12743878  
KEYWORDS 5 of 5  
SEGMENT Caldicellulosiruptor sp. Tok7B.1  
SOURCE Caldicellulosiruptor sp. Tok7B.1  
ORGANISM Bacteria; Firmicutes; Clostridia; Clostridiales;  
Syntrophomonadaceae; Caldicellulosiruptor.  
REFERENCE 1 (bases 1 to 6005)  
AUTHORS Gibbs, M.D., Reeves, R.A., Farrington, G.K., Anderson, P.,  
Williams, D.P. and Bergquist, P.L.  
TITLE Multidomain and multifunctional glycosyl hydrolases from the  
extreme thermophile caldicellulosiruptor isolate Tok7B.1  
JOURNAL Curr. Microbiol. 40 (5), 333-340 (2000)  
MEDLINE 20171169  
PUBMED 10706665  
REFERENCE 2 (bases 1 to 6005)  
AUTHORS Gibbs, M.D., Reeves, R.A., Farrington, G.K., Anderson, P.,  
Williams, D.P. and Bergquist, P.L.  
TITLE Direct Submission  
JOURNAL Submitted (27-OCT-1999) Biological Sciences, Macquarie University,  
Sydney, NSW 2109, Australia

## FEATURES

source  
1. 6005  
/organism="Caldicellulosiruptor sp. Tok7B.1"  
/mol\_type="genomic DNA"  
/isolate="Tok7B.1"  
/db\_xref="taxon:80339"  
<1..506  
/note="ORF6; multidomain"  
/codon\_start=3  
/transl\_table=11  
/product="glycosyl hydrolase 6"  
/protein\_id="AAK06393.1"  
/db\_xref="GI:12743884"  
/translation="GQPDVWNGTYTGNPNLHVKIVDYGTDLGITASLANALLYSAAT  
KEYGVSDAAKNLAKLELDRWNLYRDKLSAPKKRQYKRFQEVYIPAGWTGKM  
PNGDIVKSGVKFIDIRSKYKQDPDQKLVSNAYNAGEAPEFRHYRFAQCDAIANATY  
EILFGNQ"  
misc\_feature <1..503  
/note="Region: family 48 glycosyl hydrolase domain"







```
Qy 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 2842 ATAGGGCAAGCATGTGACATCAATTTTGTGAAGCTTAGCACGGGAGTGGAGCG 2901
Qy 84 AspThrTyrLeuGln 88
Db 2902 GATTATTACCTGGAG 2916
```

```
Qy 2 SerGlyGlyValLysValGlnTyrIlysAsnAspSerAlaProGlyAspAsnGlnIle 21
Db 1449 AGCGGACAGATAAAGGTACTGTATGCTAACAAGGACACAAACACGACGACGATA 1508
Qy 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
Db 1509 AGGCCATGGTTGAAGTAGTGAACACGGAAGCAGCATAGATTGTGACGCGGTAAACG 1568
Qy 42 ValArgTyrTyrPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTyr 61
Db 1569 ATAAGGTACTGTGTACACGGTGGATGGAGACAGGACAGAGTGGGTATCA---GATTGG 1625
Qy 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81
Db 1626 GCACAGATAGGAGCAAGCAATGTAAACATTCAAGTTTGTGAAGCTGAGCAGTAAAGT 1685
Qy 82 ThrAlaAspThrTyrLeuGln 88
Db 1686 GGAGCGGACTATTACTTTGGAG 1706
```

Search completed: October 5, 2005, 16:33:45  
Job time : 1212.63 secs

```
Qy 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 2842 ATAGGGCAAGCATGTGACATCAATTTTGTGAAGCTTAGCACGGGAGTGGAGCG 2901
Qy 84 AspThrTyrLeuGln 88
Db 2902 GATTATTACCTGGAG 2916
```

RESULT 15  
LOCUS AB008029 4743 bp DNA linear BCT 22-OCT-1997  
DEFINITION Thermophilic anaerobe NA10 gene for beta-glucanase, complete cds.  
ACCESSION AB008029  
VERSION AB008029.1 GI:2564014  
KEYWORDS beta-glucanase.  
SOURCE thermophilic anaerobe NA10  
ORGANISM thermophilic anaerobe NA10  
Bacteria.  
REFERENCE 1 (sites)  
Miyake,K., Machida,Y., Hattori,K. and Iijima,S.  
TITLE Characterization of a multi-domain cellulase from an extremely  
thermophilic anaerobe strain NA10  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 4743)  
Miyake,K., Machida,Y. and Iijima,S.  
AUTHORS Direct Submission  
TITLE Submitted (09-OCT-1997) Katsuhide Miyake, Nagoya University,  
Department of Biotechnology, School of Engineering,; Furo-cho,  
Chikusa-Ku, Nagoya, Aichi 464-01, Japan  
JOURNAL (E-mail:miyake@proc.nubio.nagoya-u.ac.jp, Tel:+81-052-789-4278,  
Fax:+81-052-789-3221)

FEATURES  
source  
1. 4743  
/organism="thermophilic anaerobe NA10"  
/mol\_type="genomic DNA"  
/strain="NA10"  
/db\_xref="taxon:67756"  
336..3338  
/codon\_start=1  
/transl\_table=11  
/product="beta-glucanase"  
/protein\_id="BA22939.1"  
/db\_xref="GI:2564015"

CDS  
/translation="MEKKLILKHFNSITPENIEIPSELLINENTVNFKASDFVNFAL  
SNNTIRGHTLVHMQQDPDFKDSNGNTLSKDALSLKQIYDVVGRYKGVYAWD  
VVNEALDRGQSDGYRRSNWYICGPEYIEKAFIWAHEADPNAKLFYNDINTENSQRQ  
FYNLKSLKEGVPIHGVGLQCHINLWPSISEIENTIKLFSSIPGIEIHITELDM  
FYQGSSTNYAVPRELLIKQAORYKELFDLFKKYKNVTVNFWGLKDDYSWLSKNY  
GKKDYPLLDENYSKYAFWSLIDPSIVFTTITLTPPEIQPLITPTPTASVTPTPT  
VTPTPTVTPT  
PWLKVNTGSSSIDLSRVITRYWTVDDGKAQSAVSDWAQIGASNVTFKVKLSSSVS  
GADYYLEIGFSGAGLQPGDGTGIBQIRFNKSDMSYNGNDMSWISQMTSYGENMK  
VTAYDGLVNGQEPGTATAPIATPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT  
TATPAPT  
GWSNLSKSTLIABIANRFLNRVPISEALILNWSQGIYKPNINITYNPELEGKNSL  
EVFDIVVQCKEVLKIMLDISIKTDAMGHYPVWYDEKFTPEDFYKACEMITNRYK  
NDDTIIADLPKHPKQWDTTFAKWDNSTDINNWKYAAETCAKRILININENLLIIV  
EGIRAYPKDDVTWTSKSSDYSTWNGNLGRVKYPINLGYQNVVSPHDYGPSV  
YQOPWFYGFPTKESLQDCWRPNWAYIMEENIAPLLIGEGSHLDGADNEKMKYLRD  
YIIENHIIHFWCFNANSQDGLGVGYDFTTDEKKYSFLKPALWQDSQGRFVGLDHK  
RPLGTNGKNIINITYNNNEPEVPATK"

ORIGIN  
Alignment Scores:  
Pred. No.: 2.59e-13 Length: 4743  
Score: 192.50 Matches: 38  
Percent Similarity: 63.22% Conservative: 17  
Best Local Similarity: 43.68% Mismatches: 31  
Query Match: 40.96% Indels: 1  
DB: 1 Gaps: 1

US-09-917-376-4 (1-89) x AB008029 (1-4743)



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model  
Run on: October 5, 2005, 06:17:10 ; Search time 136.919 Seconds  
(without alignments)  
3847.935 Million cell updates/sec

Title: US-09-917-376-4  
Perfect score: 470  
Sequence: 1 VSGGVKQVQKNDSAPGDNO.....RASFGSNPATPTADTYLQX 89

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues  
Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-O=/cgm2.1/USPTO.spool/US0917376/runat\_04102005\_164327\_26889/app\_query.fasta\_1.2524  
-DB=N\_Geneseq\_16Dec04 -QWMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US0917376 @CGN 1 1867 @runat\_04102005\_164327\_26889 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_16Dec04:\*  
1: geneseq1980s:\*  
2: geneseq1990s:\*  
3: geneseq2000s:\*  
4: geneseq2001as:\*  
5: geneseq2001bs:\*  
6: geneseq2002as:\*  
7: geneseq2002bs:\*  
8: geneseq2003as:\*  
9: geneseq2003bs:\*  
10: geneseq2003cs:\*  
11: geneseq2003ds:\*  
12: geneseq2004as:\*  
13: geneseq2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Length DB ID	Description
1	469	99.8	2289	8 ABZ77633 Nucleotid
2	469	99.8	2289	10 ADJ38292 A. cellul
3	466	99.1	3687	8 ABZ77634 Nucleotid
4	466	99.1	3687	12 ADH36637 DNA encod
5	462	98.3	2869	10 ADD22922 Acidothe

6	462	98.3	2869	10	ABZ77632	Abz77632 Nucleotid
7	462	98.3	2869	12	ADOS2314	Ados2314 Acidothe
8	436	92.8	3365	8	ABZ76162	Abz76162 A. cellul
9	409	87.0	3364	10	ADF75896	Adf75896 Acidothe
10	231.5	49.3	2600	2	AAQ15178	Aaq15178 Portion o
11	195.5	41.6	2029	2	AAZ55660	Aaz55660 DNA seque
12	195.5	41.6	2029	6	AAZ26568	Aaz26568 Active ce
13	194.5	41.4	6415	2	AAZ55662	Aaz55662 DNA seque
14	194.5	41.4	6416	6	AAZ26526	Aaz26526 Active ce
15	192.5	41.0	11706	2	AAZ55661	Aaz55661 DNA seque
16	192.5	41.0	11707	6	AAZ26525	Aaz26525 Active ce
17	190	40.4	1527	13	ADT48400	Adt48400 Bacterial
18	190	40.4	2177	11	ADOS5906	Ados5906 Bacillus
19	188	40.0	1434	6	AAZ41028	Aaz41028 CMCCase ge
20	188	40.0	1488	6	AAZ41028	Aaz41028 CMCCase ge
21	188	40.0	2510	6	ABK53202	Abk53202 Bacillus
22	186	39.6	2175	2	AAQ49820	Aaq49820 NK-1 cell
23	182.5	38.8	2977	2	AAQ13001	Aaq13001 Endol gen
24	165	35.1	501	10	ADG14263	Adg14263 Cellulose
25	165	35.1	717	10	ADG14257	Adg14257 Cellulose
26	165	35.1	1438	3	AAZ45336	Aaz45336 DNA encod
27	165	35.1	1482	2	AAZ90978	Aaz90978 DNA encod
28	165	35.1	1482	2	AAZ31562	Aaz31562 Pectate 1
29	165	35.1	5562	2	AAT86625	Aat86625 C. thermo
30	163	34.7	1500	12	ADJ35111	Adj35111 DNA encod
31	157	33.4	1314	6	ABK73393	Abk73393 Bacillus
32	157	33.4	2602	12	ADG32259	Adg32259 DNA encod
33	155.5	33.1	4059	5	AAH75059	Aah75059 Nucleotid
34	150.5	32.0	4056	5	AAH75060	Aah75060 Nucleotid
35	130.5	27.8	4452	8	ACA27429	Aca27429 Prokaryot
36	126.5	26.9	1303	10	ADC27473	Adc27473 Fusion pr
37	126.5	26.9	1747	10	ADC27474	Adc27474 Fusion pr
38	126.5	26.9	3489	10	ADC27475	Adc27475 Fusion pr
39	120.5	25.6	1775	2	AAQ13003	Aaq13003 Endo3 gen
40	114	24.3	486	2	AAQ72917	Aaq72917 Cellulose
41	114	24.3	486	2	AAV74072	Aav74072 C. cellul
42	114	24.3	499	2	AAZ24930	Aaz24930 Clostridi
43	114	24.3	507	5	AAZ11042	Aaz11042 Clostridi
44	114	24.3	768	4	AAZ86248	Aaz86248 DNA seque
45	114	24.3	768	6	ABK52403	Abk52403 E. coli c

# ALIGNMENTS

RESULT 1  
ABZ77633  
ID ABZ77633 standard; DNA; 2289 BP.  
XX  
AC ABZ77633;  
XX  
DT 03-JUN-2003 (first entry)  
XX  
DE Nucleotide sequence of the ManA polypeptide.  
XX  
KW ManA; mannanase A; glycoside hydrolase; enzyme; hemicellulose; sugar;  
XX food; feed; paper pulp; biofuel; manase; gene; ss.  
XX OS Acidothermus cellulolyticus.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..2289  
FT /tag= a  
FT /product= "ManA"  
XX  
XX WO2003012110-A1.  
XX  
XX PD 13-FEB-2003.  
XX  
XX PF 28-JUL-2001; 2001WO-US023819.  
XX  
XX PR 28-JUL-2001; 2001WO-US023819.  
XX  
XX (MIDE ) MIDWEST RES INST.  
XX



XX Ding S, Adney WS, Vinzant TB, Himmel ME;  
 XX WPI; 2003-248182/24.  
 DR P-PSDB; ABP73022.  
 XX Novel thermal tolerant mannanase A polypeptide derived from Acidothermus  
 PT cellulolyticus, useful for reducing hemicellulose in a starting material,  
 PT for processing of food, and as bulking agents in food stuffs.  
 XX Example 1; Page 23; 46pp; English.  
 XX The present sequence encodes ManA, a thermostable mannanase A polypeptide  
 CC derived from Acidothermus cellulolyticus. ManA is a member of the  
 CC glycoside hydrolase family of enzymes. ManA is useful for reducing  
 CC hemicellulose in a starting material to simpler carbohydrate units, and  
 CC ultimately to sugars which are useful in the food, feed, paper pulp, and  
 CC biofuels industries. It is useful for the processing of food and in food  
 CC stuffs as bulking agents, and for the degradation of mannan. ManA is  
 CC also useful to raise polyclonal and monoclonal antibodies that are useful  
 CC in purifying ManA, or detecting ManA polypeptide expression, and as well  
 CC as reagent tools for characterizing the molecular actions of ManA  
 CC polypeptides  
 XX SQ Sequence 2289 BP; 463 A; 700 C; 688 G; 438 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 5.36e-48 Length: 2289  
 Score: 469.00 Matches: 88  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 99.79% Indels: 0  
 DB: 8 Gaps: 0  
 US-09-917-376-4 (1-89) x ABZ77633 (1-2289)  
 Qy 1 ValSerGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
 Db 1363 GTGTCGGGTGGGTGAAGTGCAGTACAGAACAAATGATTCGGCGCCGGGTGATACACG 1422  
 Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
 Db 1423 ATCAAAACCGGTCTCCAGTTGGTGAATACGGGGTTCGTGGTGGATTTGTCACGGTG 1482  
 Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
 Db 1483 ACGGTGCGGTACTCGTTTCCACCGGGATGGTGGGTTCGTGCACACTGGTGTACACTGTGAC 1542  
 Qy 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
 Db 1543 TGGCGCGCGATGGGGTGTGGGAATATCCGCCCTCGTTCGGCTCGGTGAACCCGGCGACG 1602  
 Qy 81 ProThrAlaAspThrTyrLeuGln 88  
 Db 1603 CCGACGGCGGACACTACCTGCAG 1626

RESULT 2  
 ADJ38292  
 ID ADJ38292 standard; DNA; 2289 BP.  
 AC ADJ38292;  
 XX 06-MAY-2004 (first entry)  
 XX A. cellulolyticus mannanase ManA DNA.  
 XX mannanase A; catalytic domain GH5; carbohydrate binding domain III;  
 KW carbohydrate binding domain II; food processing; foodstuff;  
 KW bulking agent; ManA; ds; gene.  
 XX Acidothermus cellulolyticus.  
 XX Key Location/Qualifiers

CDS 1..2289  
 FT /\*tag= a  
 FT /EC number= "3.2.1.78"  
 FT /product= "Mannanase A"  
 XX US2003119093-A1.  
 XX 26-JUN-2003.  
 XX 28-JUL-2001; 2001US-00917378.  
 XX 28-JUL-2001; 2001US-00917378.  
 XX (DING/) DING S.  
 PA (ADNE/) ADNEY W S.  
 PA (VINZ/) VINZANT T B.  
 PA (HIMM/) HIMMEL M E.  
 XX Ding S, Adney WS, Vinzant TB, Himmel ME;  
 XX WPI; 2003-863435/80.  
 DR P-PSDB; ADJ38291.  
 XX New composition comprising a purified mannanase A peptide, comprising a  
 PT catalytic domain GH5, carbohydrate binding domain III or carbohydrate  
 PT binding domain II, useful in food processing and in food stuffs as  
 PT bulking agents.  
 XX Claim 11; SEQ ID NO 2; 18pp; English.  
 PS The invention relates to a composition comprising a purified mannanase A  
 CC peptide, having a catalytic domain GH5, carbohydrate binding domain III  
 CC or carbohydrate binding domain II. The composition is useful in food  
 CC processing and in foodstuffs as bulking agents. The present sequence  
 CC represents the A. cellulolyticus mannanase ManA DNA.  
 XX SQ Sequence 2289 BP; 463 A; 700 C; 688 G; 438 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 5.36e-48 Length: 2289  
 Score: 469.00 Matches: 88  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 99.79% Indels: 0  
 DB: 10 Gaps: 0  
 US-09-917-376-4 (1-89) x ADJ38292 (1-2289)  
 Qy 1 ValSerGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
 Db 1363 GTGTCGGGTGGGTGAAGTGCAGTACAGAACAAATGATTCGGCGCCGGGTGATACACG 1422  
 Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
 Db 1423 ATCAAAACCGGTCTCCAGTTGGTGAATACGGGGTTCGTGGTGGATTTGTCACGGTG 1482  
 Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
 Db 1483 ACGGTGCGGTACTCGTTTCCACCGGGATGGTGGGTTCGTGCACACTGGTGTACACTGTGAC 1542  
 Qy 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
 Db 1543 TGGCGCGCGATGGGGTGTGGGAATATCCGCCCTCGTTCGGCTCGGTGAACCCGGCGACG 1602  
 Qy 81 ProThrAlaAspThrTyrLeuGln 88  
 Db 1603 CCGACGGCGGACACTACCTGCAG 1626  
 RESULT 3  
 ABZ77634  
 ID ABZ77634 standard; DNA; 3687 BP.  
 XX ABZ77634;  
 AC ABZ77634;







DB: 12 Gaps: 0

US-09-917-376-4 (1-89) x ADH36637 (1-3687)

QY 1 ValSerGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
| | | | | : | | | | |  
Db 1750 GTGTGGGTGGCTGAGGTGACGATACAGAACAAATGATTCGGCGCCGGGTGATAACACAG 1809

QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
| | | | | : | | | | |  
Db 1810 ATCAACCGGGTCTCCAGTTGGTGAATACCGGGTCTGTCGTGGATTTGTCCACGGTG 1869

QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
| | | | | : | | | | |  
Db 1870 ACGGTGGGTACTGTTTCCACCGGGATGTTGGTGTGTCGACACTGGTGTCAACACTGTGAC 1929

QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
| | | | | : | | | | |  
Db 1930 TGGCGCGGATGGGGTGTGGGAATATCCGCGCTCTGTTCCGCTCGTGAACCCGCGACG 1989

QY 81 ProThrAlaAspThrTyrLeuGln 88  
| | | | | : | | | | |  
Db 1990 CCGACGGCGGACACCTACCTGCAG 2013

RESULT 5

ID ADD22922 standard; DNA; 2869 BP.

XX AC ADD22922;

XX 15-JAN-2004 (first entry)

XX Acidothermus cellulolyticus avicelase AvIII DNA.

XX AvIII; cellulose reduction; agricultural biomass; municipal solid waste;  
glycoside hydrolase; avicelase; ds; gene.

XX Acidothermus cellulolyticus.

XX Key Location/Qualifiers  
CDS 1..2868  
FT /\*tag= a  
FT /product= "AvIII"

XX US2003108988-A1.

XX 12-JUN-2003.

XX 18-OCT-2002; 2002US-00155400.

XX 28-JUL-2001; 2001US-00917376.

XX (DING/) DING S.  
PA (ADNE/) ADNEY W S.  
PA (VINZ/) VINZANT T B.  
PA (HIMM/) HIMMEL M E.  
XX  
PI Ding S, Adney WS, Vinzant TB, Himmel ME;  
XX WPI; 2003-810853/76.  
DR P-PSDB; ADD22921.

XX New isolated thermal tolerant avicelase polynucleotide useful for  
PT detection of a polynucleotide encoding AvIII and for reducing cellulose  
PT in a starting material, e.g. municipal solid waste.

XX Claim 17; SEQ ID NO 2; 29pp; English.

XX The invention relates to an isolated polynucleotide molecule encoding a  
CC thermostable AvIII polypeptide. The polynucleotide is useful for  
CC detection of a polynucleotide encoding AvIII. The polynucleotide is  
CC useful for reducing cellulose in a starting material which involves  
CC administering to the starting material, e.g. agricultural biomass or

CC municipal solid waste, a polypeptide molecule of the polynucleotide. The  
CC method further comprises administering a second polypeptide molecule  
CC chosen from the glycoside hydrolase family of proteins. The present  
CC sequence represents DNA encoding the Acidothermus cellulolyticus  
CC avicelase AvIII.

XX Sequence 2869 BP; 546 A; 857 C; 900 G; 565 T; 0 U; 1 Other;

Alignment Scores: 5.3e-47 Length: 2869  
Pred. No.: 462.00 Matches: 86  
Score: 100.00% Conservative: 2  
Percent Similarity: 97.73% Mismatches: 0  
Best Local Similarity: 98.30% Indels: 0  
Query Match: 10 Gaps: 0

US-09-917-376-4 (1-89) x ADD22922 (1-2869)

QY 1 ValSerGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
| | | | | : | | | | |  
Db 2605 GTGTGGGTGGGTGAGGTGACGATTAAGAATAATGATTCGGCGCCGGGTGATAATCAG 2664

QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
| | | | | : | | | | |  
Db 2665 ATCAACCGGGTTCAGGTGAGTATACCGGGTCTGTCGTGGATTTGTCCACGGTG 2724

QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
| | | | | : | | | | |  
Db 2725 ACGGTGGGTACTGTTTCCACCGGGATGTTGGTGTGTCGACACTGGTGTCAACTGTGAC 2784

QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
| | | | | : | | | | |  
Db 2785 TGGCGCGGATGGGGTGTGGGAATATCCGCGCTCTGTTCCGCTCGTGAACCCGCGACG 2844

QY 81 ProThrAlaAspThrTyrLeuGln 88  
| | | | | : | | | | |  
Db 2845 CCGACGGCGGACACCTACCTGCAG 2868

RESULT 6

ABZ77632

ID ABZ77632 standard; DNA; 2869 BP.

XX AC ABZ77632;

XX 03-JUN-2003 (first entry)

XX Nucleotide sequence of the avicelase AvIII.

XX Avicelase; AvIII; glycoside hydrolase; enzyme; cellulase; biofuel;  
KW detergent; pulp processing; paper processing; feed processing; textile;  
KW cellulose; gene; ss.

XX Acidothermus cellulolyticus.

XX Key Location/Qualifiers  
CDS 1..2869  
FT /\*tag= a  
FT /product= "AvIII"  
FT /transl\_except= (pos:2869,aa:Xaa)  
FT /note= "Xaa is an unspecified residue"

XX WO2003012090-A2.

XX 13-FEB-2003.

XX 28-JUL-2001; 2001WO-US023818.

XX 28-JUL-2001; 2001WO-US023818.

XX (MIDE ) MIDWEST RES INST.

XX Ding S, Adney WS, Vinzant TB, Himmel ME;



XX WPI; 2003-248177/24.  
 DR P-PSDB; ABP73015.  
 XX  
 XX New thermostable Aviii peptide from Acidothermus cellulolyticus, useful  
 PT for degradation of cellulose or in generating anti-Aviii antibodies for  
 PT purifying recombinant Aviii polypeptides from genetically engineered  
 PT host cells.  
 XX  
 XX Claim 3; Page 24; 44pp; English.  
 PS  
 XX The present sequence encodes a thermostable avicelase polypeptide,  
 CC designated Aviii. Aviii is a member of the glycoside hydrolase family  
 CC of enzymes, and is a cellulase. Aviii is useful in the conversion of  
 CC biomass to biofuels and biofuel additives. It may be useful in the  
 CC production of detergents, pulp and paper processing, food and feed  
 CC processing and in textile processes. The thermostable Aviii peptide is  
 CC useful in the degradation of cellulose, and in generating specific anti-  
 CC Aviii antibodies that are useful in purifying recombinant Aviii  
 CC polypeptides from genetically engineered host cells, in detecting Aviii  
 CC polypeptide expression, as well as a reagent tool for characterizing the  
 CC molecular actions of the polypeptide. The Aviii polynucleotide is useful  
 CC as a source of probes or primers in various diagnostic assays  
 XX  
 XX Sequence 2869 BP; 546 A; 857 C; 900 G; 565 T; 0 U; 1 Other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 5.3e-47 Length: 2869  
 Score: 462.00 Matches: 86  
 Percent Similarity: 100.00% Conservative: 2  
 Best Local Similarity: 97.73% Mismatches: 0  
 Query Match: 98.30% Indels: 0  
 DB: 10 Gaps: 0

US-09-917-376-4 (1-89) x ABZ77632 (1-2869)

QY 1 ValSerGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
 DB 2605 GTGTCGGGTGGGGTGAAGTGCAGTATAGAAATATGATTCGGCGCGGTGATATACAG 2664

QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
 DB 2665 ATCAAGCCGGGTTCAGGTGTGGAATACCGGGTCGTGCGTGGATTTGTCGACGGTG 2724

QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
 DB 2725 ACGTGGCGGTACTGGTTTACCCGGGATGGTGGCTCGTGACACTGGTGACACTGTGAC 2784

QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
 DB 2785 TGGCGCGGATCGGGTGTGGGAATATCCGGCTCGTTCGGTCTGGAACCGCGGACG 2844

QY 81 ProThrAlaAspThrTyrLeuGln 88  
 DB 2845 CGGACGGCGGACACCTACCTGCAG 2869

RESULT 7  
 ADO52314  
 ID ADO52314 standard; DNA; 2869 BP.  
 XX  
 XX ADO52314;  
 XX  
 XX 15-JUL-2004 (first entry)  
 XX  
 XX Acidothermus cellulolyticus avicelase III (Aviii) DNA.  
 DE  
 XX Thermostable cellulase; avicelase III; Aviii; cellulose reduction;  
 KW biomass degradation; ethanol formation; industrial chemical;  
 KW fabric treatment; gene; ds.  
 XX  
 XX Acidothermus cellulolyticus.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FH

CDS  
 FT 1. .2868  
 FT /\*tag= b  
 FT /product= "Acidothermus cellulolyticus avicelase III"  
 FT /note= "CDS contains translational exceptions"  
 FT  
 FT /partial  
 FT /note= "No stop codon"  
 FT 1. .306  
 FT sig\_peptide /\*tag= a  
 FT 307. .2865  
 FT mat\_peptide /\*tag= c  
 FT /product= "Acidothermus cellulolyticus mature avicelase  
 FT III"  
 FT 2869  
 FT misc\_feature /\*tag= d  
 FT /note= "A string of unknown nucleic acid units"  
 FT  
 XX US2004038334-A1.  
 XX 26-FEB-2004.  
 XX 28-JUL-2001; 2001US-00917376.  
 XX 28-JUL-2001; 2001US-00917376.  
 XX (DING/) DING S.  
 XX (ADNE/) ADNEY W S.  
 XX (VINZ/) VINZANT T B.  
 XX (HIMM/) HIMMEL M E.  
 XX  
 XX Ding S, Adney WS, Vinzant TB, Himmel ME;  
 PI  
 XX WPI; 2004-203224/19.  
 XX P-PSDB; ADO52313.  
 XX  
 XX Novel thermostable Aviii polypeptide of glycoside hydrolase family and  
 PT isolated from Acidothermus cellulolyticus, useful for degrading treated  
 PT biomass into simpler forms of carbohydrate.  
 XX  
 XX Claim 17; SEQ ID NO 2; 19pp; English.  
 XX  
 XX The invention relates to a thermostable cellulase enzyme, avicelase III  
 CC (Aviii) and its nucleic acid sequence. Aviii is useful for reducing  
 CC cellulose in a starting material. A thermostable Aviii peptide is useful  
 CC for degrading treated biomass into simpler forms of carbohydrate, which  
 CC is used in the formation of ethanol or other industrial chemicals. It is  
 CC also useful for treating fabrics to remove cellulose-containing stains.  
 CC The present sequence is Acidothermus cellulolyticus Aviii DNA.  
 XX  
 XX Sequence 2869 BP; 546 A; 857 C; 899 G; 565 T; 0 U; 2 Other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 5.3e-47 Length: 2869  
 Score: 462.00 Matches: 86  
 Percent Similarity: 100.00% Conservative: 2  
 Best Local Similarity: 97.73% Mismatches: 0  
 Query Match: 98.30% Indels: 0  
 DB: 12 Gaps: 0

US-09-917-376-4 (1-89) x ADO52314 (1-2869)

QY 1 ValSerGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
 DB 2605 GTGTCGGGTGGGGTGAAGTGCAGTATAGAAATATGATTCGGCGCGGTGATATACAG 2664

QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
 DB 2665 ATCAAGCCGGGTTCAGGTGTGGAATACCGGGTCGTGCGTGGATTTGTCGACGGTG 2724

QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
 DB 2725 ACGTGGCGGTACTGGTTTACCCGGGATGGTGGCTCGTGACACTGGTGACACTGTGAC 2784

QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80



Db 2785 TGGCGCGCATCGGTGGGAATATCGCGCTCGGTGGAACCCGCGACG 2844  
 Qy 81 ProThrAlaAspThrTyrLeuGln 88  
 Db 2845 CCGACGGCGGACACCTACCTGAG 2868  
 RESULT 8  
 ABZ76162  
 ID ABZ76162 standard; DNA; 3365 BP.  
 XX  
 AC ABZ76162;  
 DT 29-MAY-2003 (first entry)  
 XX  
 DE A. cellulolyticus Gux1 protein encoding DNA.  
 XX  
 KW Gux1; thermal tolerant; exoglucanase; glycoside hydrolase; cellulose;  
 KW biofuel; detergent; pulp; paper processing; feed processing; textile;  
 KW cellulase; gene; ds.  
 XX  
 OS Acidothermus cellulolyticus.  
 XX  
 FH Key Location/Qualifiers  
 CDS 1..3365  
 FT /tag= a  
 FT /product= "Gux1"  
 FT /transl\_except= (pos: 682..683, aa: Pro)  
 FT /note= "this codon has an apparent one nucleotide  
 FT basepair deletion which alters the reading frame"  
 XX  
 PN WO2003012095-A1.  
 XX  
 PD 13-FEB-2003.  
 XX  
 PF 28-JUL-2001; 2001WO-US023820.  
 XX  
 PR 28-JUL-2001; 2001WO-US023820.  
 XX  
 PA (MIDE ) MIDWEST RES INST.  
 XX  
 PI Adney WS, Ding S, Vinzant TB, Himmel ME, Decker SR;  
 PI Lantz McCarter S;  
 PI  
 DR WPI; 2003-300494/29.  
 DR P-PSDB; ABZ71656.  
 XX  
 PT New thermal tolerant Gux1 peptide having specified amino acid sequence,  
 PT useful in the degradation of cellulose to biofuels.  
 XX  
 PS Disclosure; Page 22-23; 44pp; English.  
 XX  
 CC The invention relates to a thermal tolerant Gux1 peptide from A.  
 CC cellulolyticus. The Gux1 exoglucanase is a member of the glycoside  
 CC hydrolase family and comprises a catalytic domain GH48, carbohydrate  
 CC binding domain type III, and a carbohydrate binding domain type II. The  
 CC polypeptide is useful in the degradation of cellulose into biofuel, or  
 CC for conversion of biomass to biofuel additives. It is used in detergents,  
 CC pulp and paper processing, food and feed processing, and in textile  
 CC processing. It can also be used alone or in combination with other  
 CC cellulase or glycoside hydrolases. The novel polypeptide generates  
 CC alternative cellulase enzymes capable of assisting in the commercial-  
 CC scale processing of cellulose to sugar for use in biofuel production. The  
 CC present sequence represents a A. cellulolyticus Gux1 cellulase encoding  
 CC DNA  
 XX  
 SQ Sequence 3365 BP; 669 A; 1117 C; 1041 G; 538 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 1.11e-43 Length: 3365  
 Score: 436.00 Matches: 81  
 Percent Similarity: 97.62% Conservativity: 1  
 Best Local Similarity: 96.43% Mismatches: 2

Query Match: 92.77% Indels: 0  
 DB: 8 Gaps: 0  
 US-09-917-376-4 (1-89) x ABZ76162 (1-3365)  
 Qy 5 VallysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGlnIleLysProGly 24  
 Db 112 CTCAAAGCGCAGTATAGAACAAATGATTCGGCGCGAGTGACCAACAGATCAACCGGGT 171  
 Qy 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyr 44  
 Db 172 CTCCAGTTGGTGAATACCGGGTCGTCGGTGGATTTGTCACCGTGACGGTGGGTAC 231  
 Qy 45 TrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaMet 64  
 Db 232 TGGTTCACCGGGATGGTGGGTGTCGACACTGGTGATCAACTGTGACTGGCGGGGATG 291  
 Qy 65 GlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAlaAsp 84  
 Db 292 GGGTGTGGGAATATCCGGCCCTCGTTCGGCTCGGTGAACCCCGCGACGCCGCGGAC 351  
 Qy 85 ThrTyrLeuGln 88  
 Db 352 ACCTACCTGCAG 363  
 RESULT 9  
 ADF75896  
 ID ADF75896 standard; DNA; 3364 BP.  
 XX  
 AC ADF75896;  
 XX  
 DT 26-FEB-2004 (first entry)  
 XX  
 DE Acidothermus cellulolyticus Gux1 DNA.  
 XX  
 KW Gux1; gene; ds; thermal tolerant; cellulase; glycoside hydrolase;  
 KW exoglucanase; cellulose; sugar; biofuel production.  
 XX  
 OS Acidothermus cellulolyticus.  
 XX  
 FH Key Location/Qualifiers  
 CDS 1..3364  
 FT /tag= a  
 FT /product= "Gux1 protein"  
 FT /note= "This polynucleotide sequence contains translation  
 FT exceptions"  
 XX  
 PN US2003096342-A1.  
 XX  
 PD 22-MAY-2003.  
 XX  
 PF 28-JUL-2001; 2001US-00917384.  
 XX  
 PR 28-JUL-2001; 2001US-00917384.  
 XX  
 PA (ADNE/) ADNEY W S.  
 PA (DING/) DING S. T B.  
 PA (VINZ/) VINZANT T B.  
 PA (HIMM/) HIMMEL M E.  
 PA (DECK/) DECKER S R.  
 PA (MCCA/) MCCARTER S L.  
 XX  
 PI Adney WS, Ding S, Vinzant TB, Himmel ME, Decker SR, McCarter SL;  
 PI WPI; 2003-863404/80.  
 DR P-PSDB; ADF75895.  
 XX  
 PT Novel thermal tolerant cellulase of glycoside hydrolase family,  
 PT comprising catalytic domain, first and second carbohydrate binding  
 PT domain, isolated from Acidothermus cellulolyticus, useful in cellulose  
 PT degradation.  
 XX  
 PS Claim 21; SEQ ID NO 2; 20pp; English.







```
PF 15-SEP-1998; 98EP-00810919.
XX
XX 19-SEP-1997; 97US-00932571.
XX
XX (CLRN ) CLARIANT FINANCE BVI LTD.
XX
XX Anderson P, Bergquist PL, Daniels RM, Farrington GK, Gibbs MD;
XX Morgan H, Williams DP;
XX
XX WPI; 1999-315403/27.
XX P-PSDB; AAY13494.
XX
XX New truncated cellulase proteins, useful in detergents and for producing
XX 'stonewashed' denim.
XX
XX Disclosure; Page 41-42; 65pp; English.
XX
XX The invention relates to a recombinant cellulase active protein free of
XX proteinases of native thermophilic and alkalophilic origin, comprising
XX the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel E1/2/3,
XX Cel E6 or Cel E3/B5, or a stability region from one of the defined full-
XX length sequences, or functional equivalents. Cel B5 extends from amino
XX acid A1011 to P1424 or K1425 or N1426, and Cel B4/5 extends from amino
XX acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends from
XX amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel E1/2/3
XX extends from Y39 to G812, Cel E6 extends from amino acid V1233 to K1751
XX and the stability region extends from amino acid E482 to G635 in the
XX sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new
XX enzymes are useful in laundry detergent compositions to prevent or remove
XX staining, backstaining or graying, for use on cellulosic materials
XX including cotton-containing fabrics. They are especially useful for
XX preventing redeposition of colorant during stonewashing, and for
XX processing of textiles where cellulose breakdown is required. The new
XX truncated enzymes show reduced redeposition of dye compared to using non-
XX truncated cellulase compositions
XX
XX Sequence 2029 BP; 650 A; 329 C; 520 G; 530 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 4.59e-14 Length: 2029
Score: 195.50 Matches: 38
Percent Similarity: 60.23% Conservative: 15
Best Local Similarity: 43.18% Mismatches: 34
Query Match: 41.60% Indels: 1
DB: 2 Gaps: 1

US-09-917-376-4 (1-89) x AAX55660 (1-2029)
QY 1 ValSerGlyGlyValLeuValAsnThrGlySerSerValAlaProGlyAspAsnGln 20
DB 1 ATGGGAAGTGTGTGAAGTACTGTACAAAGAACATGACAAAGTGCAGCAGGTCT 60
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAlaProGlyAspAsn 40
DB 61 ATAGGCCGCTGCTTTAAGATAGTGAATGAGGAGCAGCAGTGTGTGCTTAGCAGGTT 120
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
DB 121 AAGATAAGATAGTGTGTACACAGTGTGATGATGACAAAGCCACAGAGTCCGGTA---TGTGAC 177
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
DB 178 TGGGCACAGATAGGGGCAAGCAATGTGACATTCAATTTTGTGAAGCTTAGCAGCGGAGTG 237
QY 81 ProThrAlaAspThrTyrLeuGln 88
DB 238 AGTGAGCGGATATTACTTGAG 261

RESULT 12
AAD26568
ID AAD26568 standard; DNA; 2029 BP.
XX
XX AAD26568;
AC AAD26568;

XX 26-MAR-2002 (first entry)
XX
XX Active cellulase hybrid protein, E3/B5 DNA.
XX
XX Active cellulase protein; alkalophilic; textile processing; proteinase;
XX detergent additive; stonewashed appearance; cotton-containing denim;
XX CelB5; thermophilic; commercial detergent; E3/B5 hybrid protein; ds.
XX
XX Unidentified.
XX
XX Key Location/Qualifiers
XX CDS 1..1851
XX FT /*tag= a
XX FT /product= "E3/B5 hybrid protein"
XX
XX US6294366-B1.
XX
XX 25-SEP-2001.
XX
XX 19-AUG-1998; 98US-00136574.
XX
XX 19-SEP-1997; 97US-00932571.
XX (CLRN ) CLARIANT FINANCE BVI LTD.
XX
XX Farrington GK, Anderson P, Bergquist P, Daniels R, Gibbs MD;
XX Morgan H, Williams DP;
XX
XX WPI; 2002-081780/11.
XX P-PSDB; AAE16325.
XX
XX New cellulase active protein, useful in textile processing or commercial
XX detergents, e.g. for improving the feel or appearance of cotton-
XX containing fabrics, is stable under conditions of alkaline pH and
XX elevated temperatures.
XX
XX Disclosure; Col 71-74; 61pp; English.
XX
XX The present invention relates to a cellulase active protein, which is
XX substantially free of proteinases of native thermophilic and
XX alkalophilic origin, where the cellulase active protein consists of the
XX CelB5 amino acid sequence. The cellulase active protein is useful for
XX treating cellulosic materials including cotton-containing fabrics, as
XX detergent additives. The cellulase active protein is also useful for
XX improving the feel and/or appearance of cotton-containing fabrics, for
XX removing surface fibers from cotton-containing knits or for imparting
XX stonewashed appearance to cotton-containing denims. The present proteins
XX are stable under condition of alkaline pH and elevated temperatures, thus
XX suitable for textile processing and in commercial detergents. The present
XX sequence is E3/B5 hybrid protein DNA
XX
XX Sequence 2029 BP; 650 A; 329 C; 520 G; 530 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 4.59e-14 Length: 2029
Score: 195.50 Matches: 38
Percent Similarity: 60.23% Conservative: 15
Best Local Similarity: 43.18% Mismatches: 34
Query Match: 41.60% Indels: 1
DB: 6 Gaps: 1

US-09-917-376-4 (1-89) x AAD26568 (1-2029)
QY 1 ValSerGlyValLeuValAsnThrGlySerValAlaProGlyAspAsnGln 20
DB 1 ATGGGAAGTGTGTGAAGTACTGTACAAAGAACATGACAAAGTGCAGCAGGTCT 60
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAlaProGlyAspAsn 40
DB 61 ATAGGCCGCTGCTTTAAGATAGTGAATGAGGAGCAGCAGTGTGTGCTTAGCAGGTT 120
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
```



Db 121 AAGTAAAGTACTGTTACACAGTGGTGTGACAAAGCCACAGAGTGGCGTA---TGTGAC 177  
:::|||||:::|||||:::|||||  
Qy 61 TrpAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
|||||:::|||||:::|||||  
Db 178 TGGGCACAGATAGGGCAAGCAATGTGACATTCAATTTGTGAAGCTTAGCAGCGGAGTG 237  
:::|||||:::|||||  
Qy 81 ProThrAlaSerThrTyrLeuGln 88  
|||||:::|||||  
Db 238 AGTGGAGCGGATTATTACCTGGAG 261  
:::|||||:::|||||  
RESULT 13  
AA55662  
ID AA55662 standard; DNA; 6415 BP.  
AC  
XX  
XX  
DT 30-JUL-1999 (first entry)  
XX  
XX DNA sequence encoding truncated cellulases.  
DE  
XX Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;  
KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;  
KW cotton-containing fabric; stonewashing; ss.  
XX  
OS Unidentified.  
XX  
XX EP921188-A2.  
PN  
XX  
XX 09-JUN-1999.  
PD  
XX  
XX 15-SEP-1998; 98EP-00810919.  
PF  
XX  
XX 19-SEP-1997; 97US-00932571.  
PR  
XX  
XX (CLRN ) CLARIANT FINANCE BVI LTD.  
PA  
XX  
XX Anderson P, Bergquist PL, Daniels RM, Farrington GK, Gibbs MD;  
PI Morgan H, Williams DP;  
PI  
XX  
XX WPI: 1999-315403/27.  
DR  
XX  
XX P-PSDB; AAY13493.  
XX  
XX New truncated cellulase proteins, useful in detergents and for producing  
PT 'stonewashed' denim.  
PT  
XX  
XX Disclosure; Page 24-25; 65pp; English.  
PS  
XX  
XX The invention relates to a recombinant cellulase active protein free of  
CC proteinases of native thermophilic and alkaliphilic origin, comprising  
CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,  
CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-  
CC length sequences, or functional equivalents. Cel B5 extends from amino  
CC acid A1011 to P1424 or K1425 or N1426, and Cel B4/5 extends from amino  
CC acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends from  
CC amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel E1/2/3  
CC extends from Y39 to G812, Cel E6 extends from amino acid V1233 to K1751  
CC and the stability region extends from amino acid E482 to G635 in the  
CC sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new  
CC enzymes are useful in laundry detergent compositions to prevent or remove  
CC staining, backstaining or graying, for use on cellulosic materials  
CC including cotton-containing fabrics. They are especially useful for  
CC preventing redeposition of colorant during stonewashing, and for  
CC processing of textiles where cellulose breakdown is required. The new  
CC truncated enzymes show reduced redeposition of dye compared to using non-  
XX truncated cellulase compositions  
XX  
SQ Sequence 6415 BP; 2067 A; 1082 C; 1689 G; 1541 T; 0 U; 36 Other;

Alignment Scores:  
Pred. No.: 2,66e-13 Length: 6415  
Score: 194.50 Matches: 38  
Percent Similarity: 61.18% Conservative: 14

Best Local Similarity: 44.71% Mismatches: 32  
Query Match: 41.38% Indels: 1  
DB: 2 Gaps: 1  
US-09-917-376-4 (1-89) x AA55662 (1-6415)  
Qy 4 GlyValIysValGlnTyrIysAsnIleArgAlaSerPheGlySerValAsnProAlaThr 23  
|||||:::|||||:::|||||  
Db 2664 GGTGTGAAGTACTGTTACAAAGCAATGAGACAGTGGCAGCACAGGTTCTTAAAGGCCG 2723  
:::|||||:::|||||  
Qy 24 GlyLeuGlnLeuValAsnThrClySerSerSerValAspLeuSerThrValThrValArg 43  
:::|||||:::|||||  
Db 2724 TGGTTTAAAGTAGTGAATGGAGGCGACAGCAGTGTGTATCTTAGCAGGGTTAAGATAGA 2783  
:::|||||:::|||||  
Qy 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63  
|||||:::|||||  
Db 2784 TACTGTACACAGTGGATGGTGCACAGCCACAGAGTGGCGTA---TGTGACTGGGCACAG 2840  
:::|||||:::|||||  
Qy 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83  
:::|||||:::|||||  
Db 2841 ATAGGGCAAGCAATGTGACATTCAATTTGTGAAGCTTAGCAGCGGAGTGGAGCGC 2900  
:::|||||:::|||||  
Qy 84 AspThrTyrLeuGln 88  
|||||:::|||||  
Db 2901 GATTATTACCTGGAG 2915  
:::|||||:::|||||  
RESULT 14  
AAD26526  
ID AAD26526 standard; DNA; 6416 BP.  
AC  
XX  
XX AAD26526;  
DT 26-MAR-2002 (first entry)  
XX  
XX  
XX Active cellulase protein, cele gene.  
KW Active cellulase protein; alkaliphilic; textile processing; proteinase;  
KW detergent additive; stonewashed appearance; cotton-containing denim;  
KW CelB5; thermophilic; commercial detergent; cele gene; ds.  
XX  
XX Unidentified.  
XX  
XX  
XX Location/Qualifiers  
FH Key 634..5889  
FT CDS /\*tag= a  
FT /\*product= "Cele protein"  
FT misc\_feature 748..2538  
FT /\*tag= c  
FT /\*product= "DNA encoding E1/2 protein"  
FT misc\_feature 748..2076  
FT /\*tag= b  
FT /\*product= "DNA encoding E1 protein"  
XX  
XX US6294366-B1.  
XX  
XX 25-SEP-2001.  
PD  
XX  
XX 19-AUG-1998; 98US-00136574.  
PF  
XX  
XX 19-SEP-1997; 97US-00932571.  
PR  
XX  
XX (CLRN ) CLARIANT FINANCE BVI LTD.  
PA  
XX  
XX Farrington GK, Anderson P, Bergquist P, Daniels R, Gibbs MD;  
PI Morgan H, Williams DP;  
PI  
XX  
XX WPI: 2002-081780/11.  
DR  
XX  
XX P-PSDB; AAE16324.  
XX  
XX New cellulase active protein, useful in textile processing or commercial  
PT detergents, e.g. for improving the feel or appearance of cotton-  
PT containing fabrics, is stable under conditions of alkaline pH and  
PT elevated temperatures.



XX Disclosure; Col 37-44; 61pp; English.

XX The present invention relates to a cellulase active protein, which is

XX substantially free of proteinases of native thermophilic and

CC alkaliphilic origin, where the cellulase active protein consists of the

CC CelB5 amino acid sequence. The cellulase active protein is useful for

CC treating cellulosic materials including cotton-containing fabrics, as

CC detergent additives. The cellulase active protein is also useful for

CC improving the feel and/or appearance of cotton-containing fabrics, for

CC removing surface fibers from cotton-containing knits or for imparting

CC stonewashed appearance to cotton-containing denims. The present proteins

CC are stable under condition of alkaline pH and elevated temperatures, thus

CC suitable for textile processing and in commercial detergents. The present

CC sequence is celE gene

XX Sequence 6416 BP; 2068 A; 1082 C; 1689 G; 1541 T; 0 U; 36 Other;

Alignment Scores:

Pred. No.:	2.66e-13	Length:	6416
Score:	194.50	Matches:	38
Percent Similarity:	61.18%	Conservative:	14
Best Local Similarity:	44.71%	Mismatches:	32
Query Match:	41.38%	Indels:	1
DB:	6	Gaps:	1

US-09-917-376-4 (1-89) x AAD26526 (1-6416)

Qy 4 GlyVallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23

Db 2665 GGTGTAAGGTACTGTACAGAACAAATGAGCAAGTGCAGCACAGGTTCTATAAGGCG 2724

Qy 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43

Db 2725 TGGTTTAAGATAGTGAATGGAGGAGCAGCAGTGTGATCTTAGCAGGTTTAAGATAAGA 2784

Qy 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTIPalaAa 63

Db 2785 TACTGGTACACAGTGGATGGTGACAGCCACAGAGTGGCGTA---TGTGACTGGGCACAG 2841

Qy 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83

Db 2842 ATAGGGGCAAGCAATGTGACATCTCAATTTTGTGAAGCTTAGCAGCGGAGTGCAGTGGAGCG 2901

Qy 84 AspThrTyrLeuGln 88

Db 2902 GATTATTACCTGGAG 2916

RESULT 15

AX55661

ID AX55661 standard; DNA; 11706 BP.

XX AX55661;

AC AX55661;

XX 30-JUL-1999 (first entry)

XX DNA sequence encoding truncated cellulases Cel B4/5 and Cel B5.

DE Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;

XX Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;

KW cotton-containing fabric; stonewashed; 88.

XX Unidentified.

OS

XX EP921188-A2.

PN 09-JUN-1999.

PD 15-SEP-1998; 98EP-00810919.

XX 19-SEP-1997; 97US-00932571.

XX (CLRN ) CLARIANT FINANCE BVI LTD.

XX Anderson P, Bergquist PL, Daniels RM, Farrington GK, Gibbs MD;

PI Morgan H, Williams DP;

XX WPI; 1999-315403/27.

DR P-PSDB; AAY13492.

XX New truncated cellulase proteins, useful in detergents and for producing

PT 'stonewashed' denim.

XX Disclosure; Page 20-23; 65pp; English.

XX The invention relates to a recombinant cellulase active protein free of

XX proteinases of native thermophilic and alkaliphilic origin, comprising

CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,

CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-

CC length sequences, or functional equivalents. Cel B5 extends from amino

CC acid A1011 to P1424 or N1425 or N1426, and Cel B4/5 extends from amino

CC acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends from

CC amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel E1/2/3

CC extends from Y39 to G812, Cel E6 extends from amino acid V1233 to K1751

CC and the stability region extends from amino acid B482 to G635 in the

CC sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new

CC enzymes are useful in laundry detergent compositions to prevent or remove

CC staining, backstaining or graying, for use on cellulosic materials

CC including cotton-containing fabrics. They are especially useful for

CC preventing redeposition of colorant during stonewashing, and for

CC processing of textiles where cellulose breakdown is required. The new

CC truncated enzymes show reduced redeposition of dye compared to using non-

CC truncated cellulase compositions

XX Sequence 11706 BP; 3828 A; 1994 C; 2994 G; 2890 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.02e-12	Length:	11706
Score:	192.50	Matches:	37
Percent Similarity:	62.35%	Conservative:	16
Best Local Similarity:	43.53%	Mismatches:	31
Query Match:	40.96%	Indels:	1
DB:	2	Gaps:	1

US-09-917-376-4 (1-89) x AAX55661 (1-11706)

Qy 4 GlyVallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23

Db 4038 GGTTTGAAGGTACTATACAGAACAAATGAGCAAGTGCAGCACAGGTTCTATAAGGCGG 4097

Qy 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43

Db 4098 TGGTTTAAGATAGTGAATGGAGGAGCAGCAGTGTGATCTTAGCAGGTTAAGATAAGA 4157

Qy 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAa 63

Db 4158 TACTGGTACACAGTGGATGGTGACAAAGCCACAGAGTGGCGTA---TGTGACTGGGCACAG 4214

Qy 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83

Db 4215 ATAGGGGCAAGCAATGTGACATCTCAATTTTGTGAAGCTTAGCAGCGGAGTGCAGGCG 4274

Qy 84 AspThrTyrLeuGln 88

Db 4275 GATTATTACTTGGAG 4289

Search completed: October 5, 2005, 09:20:53

Job time : 145.919 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 5, 2005, 07:43:21 ; Search time 855.806 Seconds  
(without alignments)  
3958.515 Million cell updates/sec

Title: US-09-917-376-4  
Perfect score: 470  
Sequence: 1 VSGGVKQYKKNDSAPGDNQ.....RASFGSYNPATPTADTVLQX 89

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cpn2.1/USPTO.spool/US09917376/runat\_04102005\_164328\_26911/app\_query.fasta\_1.2524  
-DB=EST\_QPWT=fastap -SUPPIX=p2n.rst -MINMATCH=0.1 -LOOFCU=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS-human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09917376 @CNC 1.1.12009 @runat\_04102005\_164328\_26911 -NCPU=6 -ICPU=3  
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_hlc:\*  
4: gb\_est3:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_gss1:\*  
9: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	81	17.2	584	CF480372	CF480372 POL1_65 H
2	79.5	16.9	754	BQ047524	BQ047524 EST596642
3	78.5	16.7	768	CB293085	CB293085 UCRCS01_0
4	77	16.4	906	BQ722991	BQ722991 AGENCOURT
5	76	16.2	376	AA043716	AA043716 zk61f10.s
6	76	16.2	606	BE362891	BE362891 DGI_90 C1
7	76	16.2	641	CD226529	CD226529 CCI_46 E
8	76	16.2	657	BE362820	BE362820 DGI_89 C0
9	76	16.2	676	CF487649	CF487649 POL1_45_C

10	76	16.2	691	4	BG739392	BG739392 EMI_82_D0
11	76	16.2	724	8	BZ616323	BZ616323 ig58e08.g
12	76	16.2	814	4	BI218457	BI218457 602937951
13	76	16.2	905	8	BZ545807	BZ545807 OGALR72TM
14	76	16.2	934	8	CG270741	CG270741 OGWIB333TH
15	76	16.2	1001	9	CG270748	CG270748 UI-H-BWL-
16	75.5	16.1	657	2	BF513128	BF513128 UI-H-BWL-
17	75.5	16.1	1041	9	CL974250	CL974250 OSLFCC025
18	75.5	16.1	1058	7	CK231562	CK231562 ILLUMIGEN
19	75.5	16.1	2091	3	HSM800308	AL050295 Homo sapi
20	75	16.0	706	7	CK352071	CK352071 hggfha47H
21	74.5	15.9	465	2	AW491464	AW491464 UI-M-BH3-
22	74.5	15.9	613	6	CA595709	CA595709 wpalc.pk0
23	74.5	15.9	650	8	BZ894843	BZ894843 Hg4_0131
24	74	15.7	302	7	CF076729	CF076729 QHK12122.
25	74	15.7	463	9	BX959940	BX959940 Reverse 8
26	74	15.7	553	4	BI664037	BI664037 603289338
27	73.5	15.6	261	2	BF875908	BF875908 RC3-ET013
28	73.5	15.6	752	7	CN187748	CN187748 UCRCS05-0
29	73.5	15.6	798	7	CN187749	CN187749 UCRCS05-0
30	73.5	15.6	816	6	CB293980	CB293980 UCRCS01_0
31	73.5	15.6	819	6	CB293984	CB293984 UCRCS01_0
32	73.5	15.6	894	7	CK338508	CK338508 Ratus no
33	73.5	15.6	906	3	AY539888	AY539888 Ratus no
34	73.5	15.6	1039	4	BG334904	BG334904 602461219
35	73.5	15.6	1237	5	BQ065850	BQ065850 AGENCOURT
36	73	15.5	357	7	CF078232	CF078232 QHK1J23.Y
37	73	15.5	360	6	C64738	C64738 C64738 Yui.i
38	73	15.5	593	7	CF622108	CF622108 laf13d06.
39	73	15.5	776	9	BX209426	BX209426 Dario rer
40	73	15.5	861	8	BH319803	BH319803 CH230-115
41	73	15.5	894	9	CG077853	CG077853 PUFKJ89TB
42	73	15.5	924	9	CG077856	CG077856 PUFKJ89TD
43	72.5	15.4	380	4	BG816375	BG816375 daq35b08.
44	72.5	15.4	450	1	AL799805	AL799805 AL799805
45	72.5	15.4	533	2	AW963120	AW963120 EST375193

## ALIGNMENTS

CF480372 584 bp mRNA linear EST 08-SEP-2003  
POL1\_65\_H07\_b1\_A002 Pollen Sorghum bicolor cdna clone  
POL1\_65\_H07\_A002 3', mRNA sequence.  
CF480372  
CF480372.1 GI:34509241  
EST.  
Sorghum bicolor (sorghum)  
Sorghum bicolor  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD  
clade; Panicoideae; Andropogoneae; Sorghum.  
1 (bases 1 to 584)  
Cordonnier-Pratt, M.-M., Suzuk, Y., Sugano, S., Klein, R.R., Liang, C.,  
Sun, F., Sullivan, R., Eastman, A., Cannon, R., Kern, B., Morgan, J.,  
Lucas, A., Al-Sheikh, A., Jones, V., Adibi, N., Owen, A., Gao, J. and  
Pratt, L.H.  
EST database from Sorghum: pollen  
Unpublished (2003)  
Other ESTs: POL1\_65\_H07\_g1\_A002  
Contact: Cordonnier-Pratt MM  
Laboratory for Genomics and Bioinformatics  
The University of Georgia, Department of Plant Biology  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 583 0210  
Email: mmpratt@uga.edu  
Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in  
the Human Genome Center, University of Tokyo Institute of Medical  
Science; plant material and RNA prepared at Texas A & M University;  
sequencing done in the Laboratory for Genomics and Bioinformatics,  
University of Georgia. Sequence ends have been trimmed to exclude







```

Db      449 GAA-----CCTCCTCCGAGATAGTTACCTT 475
RESULT 3
LOCUS   CB293085
DEFINITION
UCRCS01_05cb12_g1 Washington Navel orange cold acclimated flavedo &
albedo cDNA library Citrus sinensis cDNA clone UCRCS01_05cb12, mRNA
sequence.
CB293085
CB293085.1 GI:28618542
EST.
Citrus sinensis
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Sapindales; Rutaceae; Citrus.
1 (bases 1 to 768)
Close,T.J., Collura,K., Fenton,R.D., Feuerbacher,O., Kim,H.R.,
Kudrna,D., Wanmaker,S., Wing,R. and Yu,Y.
Development of EST Resources and New Genetic Markers for California
Citrus
Unpublished (2003)
Contact: Timothy Close
Department of Botany & Plant Sciences, University of California
Riverside, CA, 92521-0124
Tel: 9097873318
Fax: 9097874437
Email: timothy.close@ucr.edu
Seq primer: T3.
FEATURES             Location/Qualifiers
     source
     1..768
         /organism="Citrus sinensis"
         /mol_type="mRNA"
         /cultivar="Washington navel"
         /db_xref="taxon:2711"
         /clone="UCRCS01_05cb12"
         /tissue_type="rind containing flavedo and albedo"
         /dev_stage="Mature fruit"
         /lab_host="E. coli TUC121"
         /clone_lib="Washington Navel orange cold acclimated
         flavedo & albedo cDNA library"
         /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
         Site 1: EcoRI; Site 2: XhoI; Plants were grown in the
         field at University of California, Riverside Agricultural
         Operations since 1983. The scion was Washington Navel
         orange and the rootstock Carizzo Citrange. Tissue from
         mature fruit was harvested at mid-day in January 2002
         during a cold spell, when pre-dawn temperatures were
         approximately -2 to 2 degree C. Approximately 2 cm median
         sections of the rind were excised in the field from
         several fruits, then wrapped in aluminum foil and frozen
         quickly in dry ice. Total RNA was extracted using a phenol
         extraction procedure described in J. Japanese Soc. Hort.
         Sci. (1996) 64:809-814. Poly(A) RNA was purified, a cDNA
         library was made, and 1 million primary lambda cDNA clones
         were in vivo excised to give a population of pBluescript
         SK(-) phagemids. All steps to this point were performed in
         the TJ Close lab at the University of California,
         Riverside (Fenton). Phagemids were plated, plasmid DNA
         purified, cDNA clones archived, and DNA sequences
         determined bi-directionally using an ABI350 at the
         Arizona Genomics Institute, University of Arizona
         (Collura, Feuerbacher, Kim, Kudrna, Wing, Yu).
         Chromatogram files were transmitted to UC Riverside (by
         Yu), then processed at UC Riverside (by Wanmaker) using
         the HarVest pipeline (http://harvest.ucr.edu) to remove
         vector and cloning oligo sequences and various
         contaminants, and to trim to a high quality region.
         Sequences that retained a phred 17 region of at least 100
         bases were deposited to GenBank."

ORIGIN
Alignment Scores:
Pred. No.:
58.7
Length:
906

```

```

Pred. No.:
Score:
30.7
Length:
768
Matches:
20
Percent Similarity:
57.69%
Conservative:
10
Best Local Similarity:
38.46%
Mismatches:
15
Query Match:
16.70%
Indels:
7
Gaps:
2
DB:
US-09-917-376-4 (1-89) x CB293085 (1-768)
QY      3 GlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLys 22
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      612 GCGCGCATCAAC-----AAGAACACACAGCAGCAGCAATATATAAAGTAAT 662
QY      23 ProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrVal 42
      ::|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db      663 AATAATAGCAGCGTTGTACTAGTGGTAGCGTGTGGTGGTCAAGTGGCGTGACG--- 719
      ::|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
QY      43 ArgTyrTrpPheThrArgAspGlyGlySerSerThr 54
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db      720 -----TTTTCGAAGCGCGGAGGAGGAGCAGT 746
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

RESULT 4
BQ722991
LOCUS   BQ722991
DEFINITION
AGENCOURT 8103733 Lupski sympathetic_trunk Homo sapiens cDNA clone
IMAGE:619I052 5', mRNA sequence.
ACCESSION
BQ722991
VERSION
BQ722991.1 GI:21861888
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 906)
NIH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Dr. James R. Lupski
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13590 row: n column: 13
High quality sequence stop: 512.
FEATURES             Location/Qualifiers
     source
     1..906
         /organism="Homo sapiens"
         /mol_type="mRNA"
         /db_xref="taxon:9606"
         /clone="IMAGE:619I052"
         /sex="male"
         /tissue_type="sympathetic trunk"
         /dev_stage="adult, 16 yr"
         /lab_host="DH10B"
         /note="Vector: pCMV-SPORT6 (Life Technologies); Site 1:
         NotI; Site 2: SalI; cDNA made by oligo-dT priming.
         Directionally cloned using the following adaptors:
         5'-TCGACCCAGCGGCCG-3' and
         5'-GACTAGTCTTAGATCGGACGCGCCCTT(15)-3'. Size selected >
         1 kb for average insert length 1.9 kb. This is a primary
         library, non-amplified. Library constructed by Life
         Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
         College of Medicine); available through Life
         Technologies."

ORIGIN
Alignment Scores:
Pred. No.:
58.7
Length:
906

```











is 20.

Seq primer: PolyTMix  
High quality sequence start: 89  
High quality sequence stop: 651  
POLYA=No.

## FEATURES

source

## Location/Qualifiers

```
1. .657
  /organism="Sorghum bicolor"
  /mol_type="mRNA"
  /db_xref="taxon:4558"
  /clone_lib="Dark Grown 1 (DG1)"
  /note="Organ: 5-day-old dark-grown seedlings; Vector:
  lambda Zap; Site 1: XhoI; Site 2: EcoRI; The library was
  made from poly-A RNA in the cloning vector lambda Zap II.
  Clones to be sequenced were prepared by mass excision."
```

## ORIGIN

Alignment Scores:  
Pred. No.: 51.8 Length: 657  
Score: 76.00 Matches: 23  
Percent Similarity: 44.19% Conservative: 15  
Best Local Similarity: 26.74% Mismatches: 38  
Query Match: 16.17% Indels: 10  
DB: 2 Gaps: 4

US-09-917-376-4 (1-89) x BE362820 (1-657)

```
QY 5 VallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleIysProGly 24
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 204 ATAGCAATCAGCCGAAATCAAAATCAATCCGGTCCGGAGGAGCGCTGTGCGCCCGGC 263
QY 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg-- 43
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 264 -----CGCGCCGGACAGACCTTCAACATCCAAACGATCGTGCCAGGAAA 311
QY 44 ---TyrTrp---PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 312 CCATGGTGGGGATCGCAGGAGGAGGTTTGAGCAACATGGAATATATATATGATGAA 371
QY 62 AlalaMetGlyCysGlyAsnIle-----ArgAlaSerPheGlySerValAsnPro 78
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 372 GCTTGTGTGAAGTGTGGTGACGTGTTCTTTGGAGACGGTGTCTTGCTCTGTGTCGCG 431
QY 79 AlathrProThrAlaAsp 84
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 432 AGCTTGAGTCTGGTGAT 449
```

## RESULT 9

CF487649

LOCUS

CF487649 676 bp mRNA linear EST 08-SEP-2003  
POLI\_45\_C08.bl\_A002 Pollen Sorghum bicolor cDNA clone

CF487649

CF487649.1 GI:34516518

EST.

Sorghum bicolor (sorghum)

## ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Sorghum.

1 (bases 1 to 676)

Cordonnier-Pratt,M.-M., Suzuki,Y., Sugano,S., Klein,R.R., Liang,C.,  
Sun,F., Sullivan,R., Eastman,A., Cannon,R., Kern,B., Morgan,J.,  
Lucas,A., Al-Sheikh,A., Jones,V., Adibi,N., Owen,A., Gao,J. and  
Pratt,L.H.

EST database from Sorghum: pollen

## TITLE

JOURNAL

COMMENT

Other\_ESTs: POLI\_45\_C08.g1\_A002

Contact: Cordonnier-Pratt MM

Laboratory for Genomics and Bioinformatics

The University of Georgia, Department of Plant Biology

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 583 0210

Email: mmpratt@uga.edu

Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in  
the Human Genome Center, University of Tokyo Institute of Medical  
Science; plant material and RNA prepared at Texas A & M University;  
sequencing done in the laboratory for Genomics and Bioinformatics,  
University of Georgia. Sequence ends have been trimmed to exclude  
vector and regions below phred quality 16. Three-prime sequences  
are presented as their reverse complement and have been trimmed to  
exclude polyA.

Seq primer: Sug3-14 (TAGTCTAGCGCGCGGAC)

POLYA=yes.

## FEATURES

source

## Location/Qualifiers

```
1. .676
  /organism="Sorghum bicolor"
  /mol_type="mRNA"
  /cultivar="BTx623"
  /db_xref="taxon:4558"
  /clone="POLI_45_C08_A002"
  /lab_host="DH10B-T1 phage-resistant E. coli"
  /clone_lib="Pollen"
  /note="Organ: Pollen; Vector: pME18S-FL3; Site 1: XhoI;
  Site 2: XhoI; The library was prepared from polyA+ RNA
  from pollen at the late vacuolated-vacuolated stage of
  development. Pollen was harvested from greenhouse-grown
  panicles of sorghum line BTx623. Panicles were removed
  from the flag leaf prior to emergence, when no detectable
  amylase is present in pollen of male-fertile lines. This
  stage represents pollen collected from anthers about 8-14
  days prior to anthesis. Double-stranded cDNA was cloned
  unidirectionally into different DraIII sites of the
  pME18S-FL3 vector (5-prime DraIII site is CACTGTGTG,
  3-prime DraIII site is CACCATGTG). XhoI excises the cDNA
  insert."
```

## ORIGIN

Alignment Scores:  
Pred. No.: 53.7 Length: 676  
Score: 76.00 Matches: 23  
Percent Similarity: 44.19% Conservative: 15  
Best Local Similarity: 26.74% Mismatches: 38  
Query Match: 16.17% Indels: 10  
DB: 7 Gaps: 4

US-09-917-376-4 (1-89) x CF487649 (1-676)

```
QY 5 VallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleIysProGly 24
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 214 ATAGCAATCAGCCGAAATCAAAATCAATCCGGTCCGGAGGAGCGCTGTGTCGCGCGGC 273
QY 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg--- 43
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 274 -----CGCGCCGGACAGACCTTCAACATCCAAACGATCGTGCCAGGAAA 321
QY 44 ---TyrTrp---PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 322 CCATGGTGGGGATCGCAGGAGGAGGTTTGAGCAACATGGAATATATATATGATGAA 381
QY 62 AlalaMetGlyCysGlyAsnIle-----ArgAlaSerPheGlySerValAsnPro 78
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 382 GCTTGTGTGAAGTGTGGTGACGTGTTCTTTGGAGACGGTGTCTTGCTCTGTGTCGCG 441
QY 79 AlathrProThrAlaAsp 84
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 442 AGCTTGAGTCTGGTGAT 459
```

## RESULT 10

BG739392

LOCUS

DEFINITION

EMI\_82\_D05.g1\_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA

sequence.

ACCESSION

BG739392

VERSION

BG739392.1 GI:14089081



KEYWORDS	EST.
SOURCE	Sorghum bicolor (sorghum)
ORGANISM	Sorghum bicolor
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE	1 (bases 1 to 691)
AUTHORS	Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.
TITLE	An EST database from Sorghum: developing embryos
JOURNAL	Unpublished (2000)
COMMENT	Contact: Cordonnier-Pratt MM Laboratory for Genomics and Bioinformatics The University of Georgia, Department of Plant Biology Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 583 0210 Email: mmpratt@uga.edu Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence

```

FEATURES
source
seq primer: polyA
High quality sequence start: 50
High quality sequence stop: 691
POLYA=No.
Location/Qualifiers
1..691
/organism="Sorghum bicolor"
/mol_type="mRNA"
/db_xref="taxon:4558"
/clone_lib="Embryo 1 (SM1)"
/notes="Organ: Embryos germinated for 24 hr; Vector:
pBluescript II from Lambda Zap II; Site 1: XhoI; Site 2:
EcoRI; The library was made from poly-A RNA in the cloning
vector lambda Zap II. Clones to be sequenced were
prepared by mass excision."

```

Alignment Scores:		
Pred. No.:	55.3	Length: 691
Score:	76.00	Matches: 23
Percent Similarity:	44.19%	Conservative: 15
Best Local Similarity:	26.74%	Mismatches: 38
Query Match:	16.17%	Indels: 10
DP:	4	Gaps: 4

US-09-917-376-4 (1-89) x BG739392 (1-691)

Qy	5	VallysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGlnIleLysProGly	24
ph	265	ATACGACATGCCGAAAACCTAAATGAATCCGCTCCGCGAGAGAGCCCTGCTGTCGCGCCCGGCG	324

25 I.eIIIG]nI.eIIIVe]AsnThrG]vSerSerSerValAspI.eIIISerThrValThrValArg--- 43

325 -----CGCGCCGGACAGAACCTTGAAACTTCCAAACGATCGGTGCAAGGAAA 372

44 ---TyrTm---PheThrArgAsnGlyGlySerSerThrIleValTyrAsnCysAspTrp 61

323 CCATCTCTCCCGGATCTGCACGGGACGAGGGTTTGACGCAACATGGAAATATATATGATGAA 432

62 A[aa]aMetGlvCvsgGlvAsnI]e-----ArgAlaSerPheGlvSerValAsnPro 78

433 GCTTGTCTGAAAGTGTGGTGGACGTTTCTTTTGGAGACGGTGTCTTGTCTGTGGTCCG 492

79 AlaThrProThrAlaAsp 84

493 AGCTTGCAGTCCTGGTGAAT 510

RESIST 11

Accession	Length	Topology	Source
BZ616323	724 bp	DNA	GSS 16-JAN-2003
LOC15	724 bp	DNA	GSS 16-JAN-2003

DEFINITION  
19380808:gt MOS-Zinnaberg (bina methyl triacetate) and methyl genomic sequence.  
195808 5' genomic survey sequence.

ACCESSION BZ010323

VERSION  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS

TITLE	JOURNAL	COMMENT
1. The Role of the Teacher in the Classroom	Journal of Educational Research	1980, Vol. 83, No. 1, pp. 1-10
2. The Impact of Technology on Education	Journal of Educational Research	1980, Vol. 83, No. 2, pp. 11-20
3. The Importance of Parental Involvement	Journal of Educational Research	1980, Vol. 83, No. 3, pp. 21-30
4. The Effect of Teacher Expectations on Student Achievement	Journal of Educational Research	1980, Vol. 83, No. 4, pp. 31-40
5. The Role of the School in the Community	Journal of Educational Research	1980, Vol. 83, No. 5, pp. 41-50
6. The Impact of Teacher Education on Student Achievement	Journal of Educational Research	1980, Vol. 83, No. 6, pp. 51-60
7. The Importance of Teacher Evaluation	Journal of Educational Research	1980, Vol. 83, No. 7, pp. 61-70
8. The Effect of Teacher Salary on Student Achievement	Journal of Educational Research	1980, Vol. 83, No. 8, pp. 71-80
9. The Role of the School in the Community	Journal of Educational Research	1980, Vol. 83, No. 9, pp. 81-90
10. The Impact of Teacher Education on Student Achievement	Journal of Educational Research	1980, Vol. 83, No. 10, pp. 91-100

**FEATURES**  
**SOURCE**

## ORIGIN

Alignment Scores:		
Pred. No.:	58.7	Length: 724
Score:	76.00	Matches: 34
Percent Similarity:	36.28%	Conservative: 7
Best Local Similarity:	30.09%	Mismatches: 29
Query Match:	16.17%	Indels: 43
DR:	8	Gaps: 6

US-09-017-376-A (1-89) x BZ616323 (1-724)

2	Qy	SerGlyGlyValLysValGlnTyrIlyAsnAsnAspSer-----AlaPro	16
587	Db	TCCGAGCGTCAGACTCCCATACAGGAGCGCAACAGCAACGTCGCGCAGCTGCGCCC	528
17	Qy	Gly-----AspAsnGlnIleLysProGlyLeuGlnLeuValAsn-----	29
527	Db	GGCGCCCGCAGCTGAGCAGCTCGCGGAGCCACGACTGATGCTTATCAACCGTGGCGC	468
30	Qy	Thr-GlySerSerValAsp-----	36
467	Db	ACGAGGAAGTTCGCACTCCCGAGATCGTCGGAGCGGTCACAGAGCTGGGTTCGAG	408
37	Qy	-----LeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSe	53
407	Db	GTCATCCCGATCGAGCCCGGTAGACACTCAGCTGGAGGACTTCGCCCGGACCGTCGAC	348
53	Qy	rThrLeuValTyrAsnCysAspTrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPh	73
347	Db	TCGTGCG-----ACGTGCTCATGCGCGCAGCAGCGG-----	317



Qy 73 eGlySerValAsnProAlaThrProThrAlaAspThr 85  
Db 316 -----CCGGCCTCACCACACTTCTTCTTC 293

RESULT 12  
LOCUS BI218457 814 bp mRNA linear EST 11-JUL-2001  
DEFINITION 602937951F1 NCI\_CGAP\_Li9 Mus musculus cDNA clone IMAGE:5101069 5', mRNA sequence.  
ACCESSION BI218457  
VERSION BI218457.1 GI:14671901  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1. (bases 1 to 814)  
REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished (1999)  
JOURNAL  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)  
Tissue Procurement: Jeffrey E. Green, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Plate: L14M1243 row: f column: 14  
High quality sequence stop: 792.  
Location/Qualifiers  
FEATURES  
source  
1..814  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5101069"  
/lab\_host="DH10B (T1 phage-resistant)"  
/clone\_lib="NCI\_CGAP\_Li9"  
/note="Organ: liver; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.9 kb. Constructed by Life Technologies. Note: this is a NCI\_CGAP Library."

Alignment Scores:  
Pred. No.: 68.3 Length: 814  
Score: 76.00 Matches: 28  
Percent Similarity: 44.83% Conservative: 11  
Best Local Similarity: 32.18% Mismatches: 29  
Query Match: 16.17% Indels: 19  
Gaps: 4

US-09-917-376-4 (1-89) x BI218457 (1-814)

Qy 2 SerGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnle 21  
Db 33 GCAGGTGGAGCCCGAGTATGGGAGCGGAACACACAGATCGCCAGGACAATGAGCAG--- 89

Qy 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41  
Db 90 -----AGTTCCGAGTGGAGCCTCGAGACCTG--- 116

Qy 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeu-----ValTyrAsnCys 59  
Db 117 CTCGCTACTACACACAGAGCGCGGCTCTCACAGATCCAGCGATCGGCTGT 176

Qy 60 Asp-TripAlaAlaMetGlyCysGlyAsnleArgAlaSerPheGlySerValAsnProAl 79  
Db 177 GACGTGGGGTGGAGCGGCGCC-----TCCTCCGCGGGTACGAGCAGGTGCGCA 224

Qy 79 aThrProThrAlaAspThr 85

Db 225 TACGACGGCTCGGATTACA 243  
||| ||| ||| |||

RESULT 13  
LOCUS BZ545807 905 bp DNA linear GSS 16-DEC-2002  
DEFINITION OGALR72TM ZM2\_0.7\_1.5\_KB Zea mays genomic clone ZMMBMA0089L24, genomic survey sequence.  
ACCESSION BZ545807  
VERSION BZ545807.1 GI:27099699  
KEYWORDS GSS.  
SOURCE Zea mays  
ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.  
1. (bases 1 to 905)  
REFERENCE Whitehead,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Budiman,M.A., Hedell,J.A., Rohlfing,T., Citek,R.W., Numborg,A., Robbins,D. and Lakey,N.  
TITLE Consortium for Maize Genomics  
JOURNAL Unpublished (2002)  
COMMENT Other\_GSSs: OGALR72TC  
Contact: Cathy Whitelaw  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843  
Fax: 301-838-0208  
Email: [whitelaw@tigr.org](mailto:whitelaw@tigr.org)  
Seq primer: TR  
Class: sheared ends.  
Location/Qualifiers  
FEATURES  
source  
1..905  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMBMA0089L24"  
/clone\_lib="ZM2\_0.7\_1.5\_KB"  
/note="Vector: pBCKS-; Site 1: HincII; 0.7-1.5 kb methylation filtered genomic DNA library"

ORIGIN  
Alignment Scores:  
Pred. No.: 78.3 Length: 905  
Score: 76.00 Matches: 34  
Percent Similarity: 36.28% Conservative: 7  
Best Local Similarity: 30.09% Mismatches: 29  
Query Match: 16.17% Indels: 43  
Gaps: 6

US-09-917-376-4 (1-89) x BZ545807 (1-905)

Qy 2 SerGlyValLysValGlnTyrLysAsnAspSer-----AlaPro 16  
Db 87 TCCGACGGCTGAGTATCCATCCATACAGGAGGCGGACAGCAGCTGCGGCGTGGCCC 146

Qy 17 Gly-----AspAsnGlnleLysProGlyLeuGlnLeuValAsn----- 29  
Db 147 GCGCGCCCGAGCTGAGCAGCTGCGCGGAGGCCACGACTGATGCTTATCAACCGTGGCCGC 206

Qy 30 Thr-GlySerSerValAsp----- 36  
Db 207 ACGAGGAAGTTCGTCACTTCCCGAGATGCTCGGAGCGGTCCAGAGCGTGGGTTCGAG 266

Qy 37 -----LeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSe 53  
Db 267 GTCATCCCGATCGAGCGCGGTAGACCTCAGCGTGGAGGACTTCCCGCGGACCGTGCAC 326

Qy 53 rThrLeuValTyrAsnCysAspTrpAlaAlaMetGlyCysGlyAsnleArgAlaSerPh 73  
Db 327 TCGTGGC-----ACGTGCTCATGGCGCGCAGCGGG----- 357



```

Qy 73 eGlySerValAsnProAlaThrProThrAlaAspThr 85
Db 358 -----CGGGCTCACCACACTTCTTCTTCC 381

RESULT 14
CG270741/c
LOCUS CG270741 934 bp DNA linear GSS 25-AUG-2003
DEFINITION OGWIB33TV ZM 0.7 1.5 KB Zea mays genomic clone ZMMBMA0583E18,
genomic survey sequence.
ACCESSION CG270741
VERSION CG270741.1 GI:34182882
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 934)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
Consortium for Maize Genomics
TITLE Unpublished (2002)
JOURNAL Other GSSs: OGWIB33TV
COMMENT Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.
FEATURES
Location/Qualifiers
source
1..934
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMA0583E18"
/clone_lib="ZM 0.7 1.5 KB"
/notes="Vector: pBCSK; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Alignment Scores:
Pred. No.: 81.5 Length: 934
Score: 76.00 Matches: 34
Percent Similarity: 36.28% Conservative: 7
Best Local Similarity: 30.09% Mismatches: 29
Query Match: 16.17% Indels: 43
DB: 9 Gaps: 6

US-09-917-376-4 (1-89) x CG270741 (1-934)
Qy 2 SerGlyGlyValLysValGlnTyrLysAsnAspSer-----AlaPro 16
Db 804 TCGCGAGGGCTCAGCATCCATACAGAGGGGCAACAGCAACGCTCGCGAGCTGGCGCC 745
Qy 17 Gly-----AspAsnGlnIleLysProGlyLeuGlnLeuValAsn----- 29
Db 744 GGGCCCCAGCTGAGCAGCTGAGCGGAGCGGAGCCACGACTGATGCTTATCAACGTCGCCGC 685
Qy 30 Thr-GlySerSerValAsp----- 36
Db 684 ACAGGAGAGTTGCTCAACTTCCGGAGATCGTCGGAGCGGTCCAGAGGCTGGGTTCCGAG 625
Qy 37 -----LeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSe 53
Db 624 GTCATCCCGATCGAGCGCGTAGAGACCTCAGCGTGGAGGACTTCGCCCGGACCGTCGAC 565
Qy 53 rThrLeuValTyrAsnCysAspTrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPh 73
Db 564 TCGTGGG-----ACGTGCTCTGGCGCGCCGCGG----- 534

```

```

Qy 73 eGlySerValAsnProAlaThrProThrAlaAspThr 85
Db 533 -----CGGGCTCACCACACTTCTTCTTCC 510

RESULT 15
CG270748
LOCUS CG270748 1001 bp DNA linear GSS 25-AUG-2003
DEFINITION OGWIB33TV ZM 0.7 1.5 KB Zea mays genomic clone ZMMBMA0583E18,
genomic survey sequence.
ACCESSION CG270748
VERSION CG270748.1 GI:34182889
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 1001)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
Consortium for Maize Genomics
TITLE Unpublished (2002)
JOURNAL Other GSSs: OGWIB33TH
COMMENT Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
FEATURES
Location/Qualifiers
source
1..1001
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMA0583E18"
/clone_lib="ZM 0.7 1.5 KB"
/notes="Vector: pBCSK; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Alignment Scores:
Pred. No.: 89.1 Length: 1001
Score: 76.00 Matches: 34
Percent Similarity: 36.28% Conservative: 7
Best Local Similarity: 30.09% Mismatches: 29
Query Match: 16.17% Indels: 43
DB: 9 Gaps: 6

US-09-917-376-4 (1-89) x CG270748 (1-1001)
Qy 2 SerGlyGlyValLysValGlnTyrLysAsnAspSer-----AlaPro 16
Db 418 TCGCGAGGGCTCAGCATCCATACAGAGGGGCAACAGCAACGCTCGCGAGCTGGCGCC 477
Qy 17 Gly-----AspAsnGlnIleLysProGlyLeuGlnLeuValAsn----- 29
Db 478 GGGCCCCAGCTGAGCAGCTGAGCGGAGCGGAGCCACGACTGATGCTTATCAACGTCGCCGC 537
Qy 30 Thr-GlySerSerValAsp----- 36
Db 538 ACAGGAGAGTTGCTCAACTTCCGGAGATCGTCGGAGCGGTCCAGAGGCTGGGTTCCGAG 597
Qy 37 -----LeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSe 53
Db 598 GTCATCCCGATCGAGCGCGTAGAGACCTCAGCGTGGAGGACTTCGCCCGGACCGTCGAC 657
Qy 53 rThrLeuValTyrAsnCysAspTrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPh 73

```



Db 658 TCGTGG-----ACGTGCTCATGGGCGGCGCACGGGG----- 688  
Qy 73 eGlySerValAsnProAlaThrProThrAlaaspThr 85  
Db 689 -----CGGGCCTCACCAACTTCTTCTCC 712

Search completed: October 5, 2005, 21:09:32  
Job time : 860.806 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model  
Run on: October 5, 2005, 07:47:30 ; Search time 42.3629 Seconds  
(without alignments)  
3437.649 Million cell updates/sec

Title: US-09-917-376-4  
Perfect score: 470  
Sequence: 1 VSGGVKQVKNDSAPGDNQ.....RASFGSVNPATPTADTVLQX 89

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues  
Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US09917376/runat 04102005.164328 26925/app query.fasta 1.2524  
-DB=Issued Patents NA -QWTF=fastap -SUPFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09917376 @CGN 1 1 516 @runat 04102005.164328 26925 -NCPU=6 -ICPU=3  
-NO WMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:  
1: /cgn2\_6/ptodata/1/ina/5A COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	195.5	41.6	2029	3	US-09-136-574A-46 Sequence 46, Appl
2	194.5	41.4	6416	3	US-09-136-574A-2 Sequence 2, Appli
3	192.5	41.0	11707	3	US-09-136-574A-1 Sequence 1, Appli
4	182.5	38.8	2977	2	US-07-862-588B-1 Sequence 1, Appli
5	165	35.1	1438	4	US-09-339-159B-3 Sequence 3, Appli
6	165	35.1	1482	3	US-09-198-956-9 Sequence 9, Appli
7	165	35.1	1482	3	US-09-198-955A-11 Sequence 11, Appl
8	165	35.1	1482	3	US-09-694-531-11 Sequence 11, Appl
9	165	35.1	1482	3	US-09-670-141-9 Sequence 9, Appli
10	165	35.1	1482	4	US-10-072-152-11 Sequence 11, Appl
11	155.5	33.1	4059	4	US-09-784-554B-1 Sequence 1, Appli
12	150.5	32.0	4056	4	US-09-784-554B-3 Sequence 3, Appli

13	120.5	25.6	1624	2	US-07-862-588B-6 Sequence 6, Appli
14	120.5	25.6	1775	2	US-07-862-588B-5 Sequence 5, Appli
15	114	24.3	486	1	US-08-048-164A-1 Sequence 1, Appli
c 16	114	24.3	486	1	US-08-048-164A-3 Sequence 3, Appli
c 17	114	24.3	486	1	US-08-460-462-1 Sequence 1, Appli
c 18	114	24.3	486	1	US-08-460-462-3 Sequence 3, Appli
c 19	114	24.3	486	1	US-08-460-457-1 Sequence 1, Appli
c 20	114	24.3	486	1	US-08-460-457-3 Sequence 3, Appli
c 21	114	24.3	486	1	US-08-460-458-1 Sequence 1, Appli
c 22	114	24.3	486	2	US-08-460-455-1 Sequence 1, Appli
c 23	114	24.3	486	2	US-08-460-455-3 Sequence 3, Appli
c 24	114	24.3	486	2	US-08-330-394A-1 Sequence 1, Appli
c 25	114	24.3	486	2	US-08-330-394A-3 Sequence 3, Appli
c 26	114	24.3	486	3	US-09-006-636-5 Sequence 5, Appli
c 27	114	24.3	499	3	US-09-006-632-5 Sequence 5, Appli
c 28	114	24.3	499	3	US-09-325-274-5 Sequence 2, Appl
c 29	114	24.3	499	3	US-09-277-716-21 Sequence 21, Appl
c 30	113	24.0	1146	3	US-09-609-161B-21 Sequence 21, Appl
c 31	113	24.0	1146	3	US-09-335-586-3 Sequence 3, Appli
c 32	77	16.4	8107	4	US-09-949-016-12565 Sequence 12565, A
c 33	76	16.2	136058	4	US-09-949-016-17064 Sequence 17064, A
c 34	76	16.2	136480	4	US-09-949-016-17064 Sequence 333, App
c 35	75.5	16.1	5714	4	US-09-620-312D-393 Sequence 3752, Ap
c 36	72.5	15.4	819	4	US-09-902-540-3752 Sequence 1182, Ap
c 37	72.5	15.4	17897	4	US-09-526-193A-21 Sequence 21, Appl
c 38	72.5	15.4	20284	4	US-09-526-193A-21 Sequence 962, App
c 39	70	14.9	10892	4	US-09-902-540-962 Sequence 39, Appl
c 40	70	14.9	25165	3	US-09-453-702B-39 Sequence 1313, A
c 41	70	14.9	246444	4	US-09-949-016-13113 Sequence 12513, A
c 42	69.5	14.8	103987	4	US-09-949-016-17050 Sequence 17050, A
c 43	69.5	14.8	103988	4	US-09-949-016-128395 Sequence 128395, Sequence 128396,
c 44	68.5	14.6	601	4	
c 45	68.5	14.6	601	4	

ALIGNMENTS

RESULT 1  
US-09-136-574A-46  
; Sequence 46, Application US/09136574A  
; Patent No. 6294366  
; GENERAL INFORMATION:  
; APPLICANT: Farrington, Graham K.  
; Anderson, Paige  
; Gibbs, Moreland  
; Bergquist, Peter  
; Daniels, Roy  
; Morgan, Hugh W.  
; Williams, Diane P.  
; TITLE OF INVENTION: Compositions and Methods for Treating Cellulose Containing Fabrics Using Truncated Cellulase Enzyme Compositions

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: 1997US0001/CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-540-9200  
TELEFAX: 215-540-5818  
TELEX: <Unknown>  
INFORMATION FOR SEQ ID NO: 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2029 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 46:

US-09-136-574A-46

Alignment Scores:			
Pred. No.:	6,46e-17	Length:	2029
Score:	195.50	Matches:	38
Percent Similarity:	60.23%	Conservative:	15
Best Local Similarity:	43.18%	Mismatches:	34
Query Match:	41.60%	Indels:	1
DB:	3	Gaps:	1

US-09-917-376-4 (1-89) x US-09-136-574A-46 (1-2029)

Qy	1	ValSerGlyGlyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln	20
Db	1	ATGGAGTGGTGTGAGGTACTGTACAGAACAAATGAGCAAGTGGCAGCAGGTTCT	60
Qy	21	IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal	40
Db	61	ATAAGCGGTGTTAAGATAGTGAATGGAGCGCAGCAGTGTGTATCTTAGCAGGTT	120
Qy	41	ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp	60
Db	121	AAGATAAGACTGTGTACACAGTGGATGGTGCACAGCCACAGAGTGGCGTA---TGTGAC	177
Qy	61	TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr	80
Db	178	TGGGCACAGATAGGGGCAAGCAATTCATTCATTCAATTTGTGAAGCTTAGCAGCGGAGTG	237
Qy	81	ProThrAlaAspThrTyrLeuGln	88
Db	238	AGTGGAGCGGATTATTACCTGGAG	261

## RESULT 2

US-09-136-574A-2  
Sequence 2, Application US/09136574A  
Patent No. 6294366

## GENERAL INFORMATION:

APPLICANT: Farrington, Graham K.  
Anderson, Paige  
Gibbs, Moreland  
Bergquist, Peter  
Daniels, Roy  
Morgan, Hugh W.  
Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for  
Treating Cellulose Containing  
Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49

CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson  
STREET: Spring House Corporate Center, P.O. Box 457  
CITY: Spring House  
STATE: PA  
COUNTRY: USA  
ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/136,574A  
FILING DATE: 19-Aug-1998  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/932,571  
FILING DATE: September 19, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: 1997US0001/CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-540-9200  
TELEFAX: 215-540-5818  
TELEX: <Unknown>  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6416 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-136-574A-2

Alignment Scores:			
Pred. No.:	4,68e-16	Length:	6416
Score:	194.50	Matches:	38
Percent Similarity:	61.18%	Conservative:	14
Best Local Similarity:	44.71%	Mismatches:	32
Query Match:	41.38%	Indels:	1
DB:	3	Gaps:	1

US-09-917-376-4 (1-89) x US-09-136-574A-2 (1-6416)

Qy	4	GlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro	23
Db	2665	GGTGTGAGGTACTGTGTACAGAACAAATGAGCAAGTGGCAGCAGGTTCTATAAGCGCG	2724
Qy	24	GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg	43
Db	2725	TGGTTAAGTAGTAGTATGGAGCGCAGCAGTGTGTATCTTAGCAGGTTAAGATAAGA	2784
Qy	44	TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla	63
Db	2785	TACTGTACACAGTGTGTGCACAGCCACAGAGTGGCGTA---TGTGACTGGGCACAG	2841
Qy	64	MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla	83
Db	2842	ATAGGGCAAGCAATGTGACATTCATTTTGTGAAGCTTAGCAGCGGAGTGGAGCGG	2901
Qy	84	AspThrTyrLeuGln	88
Db	2902	GATTATTACCTGGAG	2916

## RESULT 3

US-09-136-574A-1

Sequence 1, Application US/09136574A  
Patent No. 6294366

## GENERAL INFORMATION:

APPLICANT: Farrington, Graham K.  
Anderson, Paige  
Gibbs, Moreland  
Bergquist, Peter  
Daniels, Roy  
Morgan, Hugh W.  
Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for

Treating Cellulose Containing  
Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49

CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson

STREET: Spring House Corporate Center, P.O. Box 457

CITY: Spring House



**us-09-917-376-4.p2n.rni**

STREET: 405 Lexington Avenue, 62nd floor  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10017  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/862,588B  
FILING DATE: 19920727  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DK 164/90  
FILING DATE: 19-JAN-1990  
PRIOR APPLICATION DATA: PCT/DK91/00013  
APPLICATION NUMBER:  
FILING DATE: 18-JAN-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Zelson, Steve T. / Lambiris, Elias J.  
REGISTRATION NUMBER: 30,335 / 33,728  
REFERENCE/DOCKET NUMBER: 3425.204-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 867 0123  
TELEFAX: 212 867 0298  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2977 base pairs

```

; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORGANISM SOURCE:
; ORGANISM: Bacillus lautus
; STRAIN: NCIMB 40250
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 677..2776
; OTHER INFORMATION:
;
US-07-862-588E-1

Alignment Scores:
Pred. No.: 7 14e-15
Score: 182.50
Percent Similarity: 66.27%
Best Local Similarity: 43.37%
Query Match: 38.83%
DR: 2
Length: 2977
Matches: 36
Conservative: 19
Mismatches: 25
Indels: 3
Gaps: 2

```

```

US-09-917-376-4 (1-89) x US-07-862-588B-1 (1-2977)
Qy      6  LysValGlnTyrLysAsnAsnSerAlaProGlyAspAsnGlnIleLysProGlyLeu  25
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     144  CGACTGCAGTACAGAGCGGGCCGATCAAAATGCAGCCGACACCAAGCCGTCCTTC  203
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy      26  GlnLeuValAsnThrGlySerSerSerValAspLeuSerThrValThrValArgTyrTrp  45
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     204  AACATCAAAACAAACGGTACTTCGGCTGTGTATTAGACGCTCAAAATCCGCTACTAC  263
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy      46  PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTTPAlaAlaMetGly  65
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     264  TTCACCAAGGATGGTTCTTGGCGGGCGTGAACGGCTGG---ATCGACTGGCGCAGCTCGGC  320
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy      66  CysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAlaAspThr  85
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     321  GGCAGCAACATTCAGATCTGTTGGC-----AACCATACTGGCAGCAATTCGATACG  374
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy      86  TyrLeuGln  88
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     375  TACGTGGAG  383
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

22



4



Alignment Scores:  
Pred. No.: 6.97e-13 Length: 1482  
Score: 165.00 Matches: 35  
Percent Similarity: 59.79% Conservativeness: 23  
Best Local Similarity: 36.08% Mismatches: 29  
Query Match: 35.11% Indels: 10  
DB: 3 Gaps: 2

US-09-917-376-4 (1-89) x US-09-198-955A-11 (1-1482)

QY 1 ValSerGlyVallyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
DB 991 GTATCAGCAATTTGAAGTTGAATCTACACAGCAATCTTCAGATACCTACTACTCA 1050  
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
DB 1051 ATCAATCCTCAGTTCAGGTTACTAATACCGAAGCAGTGCAATTTGTCCTCAACTC 1110  
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
DB 1111 ACATTGAGATATTATTATACAGTAGACGACAGAAAGATCAGACCTTCTGG---TGTGAC 1167  
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71  
DB 1168 CATGCTGCAATTAATCGCGAGTAAACGCGAGCTACACGGAATTTACTTCAATGTAAAGGA 1227  
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88  
DB 1228 ACATTGTAAATGAGTTCTCTCAACAATAACGCGAGACCTTACCTTGAA 1278

## RESULT 8

US-09-694-531-11  
; Sequence 11, Application US/09694531  
; Patent No. 6368843  
; GENERAL INFORMATION:  
; APPLICANT: Andersen, Lene N.  
; APPLICANT: Schulein, Martin  
; APPLICANT: Lange, Niels E.  
; APPLICANT: Bjornvad, Mads E.  
; APPLICANT: Moller, Soren  
; APPLICANT: Glad, Sanne O. S.  
; APPLICANT: Kauppinen, Markus S.  
; APPLICANT: Schnorr, Kirk  
; APPLICANT: Kongsbak, Lars  
; TITLE OF INVENTION: No. 6368843el Pectate Lyases  
; FILE REFERENCE: 5378.200-US  
; CURRENT FILING DATE: 2000-10-23  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/067,249  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/067,240  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 1343/97  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 1344/97  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/067,249  
; PRIOR FILING DATE: 1997-12-02  
; PRIOR APPLICATION NUMBER: 60/067,240  
; PRIOR FILING DATE: 1997-12-02  
; PRIOR APPLICATION NUMBER: 09/073,684  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: 09/184,217  
; PRIOR FILING DATE: 1998-11-02  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11  
; LENGTH: 1482  
; TYPE: DNA  
; ORGANISM: Clostridium thermoecellum  
US-09-694-531-11

Alignment Scores:  
Pred. No.: 6.97e-13 Length: 1482

Score: 165.00 Matches: 35  
Percent Similarity: 59.79% Conservativeness: 23  
Best Local Similarity: 36.08% Mismatches: 29  
Query Match: 35.11% Indels: 10  
DB: 3 Gaps: 2  
US-09-917-376-4 (1-89) x US-09-694-531-11 (1-1482)  
QY 1 ValSerGlyVallyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
DB 991 GTATCAGCAATTTGAAGTTGAATCTACACAGCAATCTTCAGATACCTACTACTCA 1050  
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
DB 1051 ATCAATCCTCAGTTCAGGTTACTAATACCGAAGCAGTGCAATTTGTCCTCAACTC 1110  
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
DB 1111 ACATTGAGATATTATTATACAGTAGACGACAGAAAGATCAGACCTTCTGG---TGTGAC 1167  
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71  
DB 1168 CATGCTGCAATTAATCGCGAGTAAACGCGAGCTACACGGAATTTACTTCAATGTAAAGGA 1227  
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88  
DB 1228 ACATTGTAAATGAGTTCTCTCAACAATAACGCGAGACCTTACCTTGAA 1278

## RESULT 9

US-09-670-141-9  
; Sequence 9, Application US/09670141  
; Patent No. 6429000  
; GENERAL INFORMATION:  
; APPLICANT: Andersen, Lene N.  
; APPLICANT: Schulein, Martin  
; APPLICANT: Lange, Niels Erik K.  
; APPLICANT: Bjornvad, Mads E.  
; APPLICANT: Schnorr, Kirk  
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus  
; FILE REFERENCE: 5377.200-US  
; CURRENT FILING DATE: 2000-09-26  
; PRIOR FILING DATE: 1998-11-24  
; PRIOR APPLICATION NUMBER: 1344/97  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/067,240  
; PRIOR FILING DATE: 1997-12-02  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 9  
; LENGTH: 1482  
; TYPE: DNA  
; ORGANISM: Bacillus licheniformis  
US-09-670-141-9

Alignment Scores:  
Pred. No.: 6.97e-13 Length: 1482  
Score: 165.00 Matches: 35  
Percent Similarity: 59.79% Conservativeness: 23  
Best Local Similarity: 36.08% Mismatches: 29  
Query Match: 35.11% Indels: 10  
DB: 3 Gaps: 2

US-09-917-376-4 (1-89) x US-09-670-141-9 (1-1482)

QY 1 ValSerGlyVallyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
DB 991 GTATCAGCAATTTGAAGTTGAATCTACACAGCAATCTTCAGATACCTACTACTCA 1050  
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40



Db 1051 ATCAATCCTCAGTTCAAGGTTACTAATACCGAAGCAGTGAATGATTTGTCCTCAAACTC 1110  
Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
Db 1111 ACATTGAGATATTATTATACAGTACGAGCAGAAAGATCAGACCTTCTGG---TGTGAC 1167  
Qy 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71  
Db 1168 CATGCTGCAATAATCGGCAGTAAACGGCAGCTACAACGGAATTTCTCAAAATGTTAAAGGA 1227  
Qy 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88  
Db 1228 ACATTGTAAAAATGAGTTCTCTCAACAAATAACGAGCAGACCTACCTTTGAA 1278

## RESULT 10

US-10-072-152-11  
; Sequence 11, Application US/10072152  
; Patent No. 6677147  
; GENERAL INFORMATION:  
; APPLICANT: Andersen, Lene N.  
; APPLICANT: Schulein, Martin  
; APPLICANT: Lange, Niels E.  
; APPLICANT: Bjornvad, Mads E.  
; APPLICANT: Moller, Soren  
; APPLICANT: Glad, Samne O. S.  
; APPLICANT: Kauppinen, Markus S.  
; APPLICANT: Schnorr, Kirk  
; APPLICANT: Kongsbak, Lars  
; TITLE OF INVENTION: No. 6677147el Pectate Lyases  
; FILE REFERENCE: 5378.200-US  
; CURRENT APPLICATION NUMBER: US/10/072,152  
; CURRENT FILING DATE: 2002-02-07  
; PRIOR APPLICATION NUMBER: US/09/198,955  
; PRIOR FILING DATE: 1998-11-24  
; PRIOR APPLICATION NUMBER: 1343/97  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 1344/97  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/067,249  
; PRIOR FILING DATE: 1997-12-02  
; PRIOR APPLICATION NUMBER: 60/067,240  
; PRIOR FILING DATE: 1997-12-02  
; PRIOR APPLICATION NUMBER: 09/073,684  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: 09/184,217  
; PRIOR FILING DATE: 1998-11-02  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11  
; LENGTH: 1482  
; TYPE: DNA  
; ORGANISM: Clostridium thermocellum  
US-10-072-152-11

Alignment Scores:  
Pred. No.: 6,97e-13 Length: 1482  
Score: 165.00 Matches: 35  
Percent Similarity: 59.79% Conservative: 23  
Best Local Similarity: 36.08% Mismatches: 29  
Query Match: 35.11% Indels: 10  
DB: 4 Gaps: 2

US-09-917-376-4 (1-89) x US-10-072-152-11 (1-1482)

Qy 1 ValSerGlyValValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
Db 991 GTATCAGCAATTGAGGTTGAATCTTCAACAGCAATCTTCAGATACTAACTCA 1050  
Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
Db 1051 ATCAATCCTCAGTTCAAGGTTACTAATAACCGAAGCAGTGAATGATTTGTCCTCAAACTC 1110  
Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60

Db 1111 ACATTGAGATATTATTATACAGTACGAGCAGAAAGATCAGACCTTCTGG---TGTGAC 1167  
Qy 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71  
Db 1168 CATGCTGCAATAATCGGCAGTAAACGGCAGCTACAACGGAATTTCTCAAAATGTTAAAGGA 1227  
Qy 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88  
Db 1228 ACATTGTAAAAATGAGTTCTCTCAACAAATAACGAGCAGACCTACCTTTGAA 1278

## RESULT 11

US-09-784-554B-1  
; Sequence 1, Application US/09784554B  
; Patent No. 6815192  
; GENERAL INFORMATION:  
; APPLICANT: Schnorr, Kirk  
; APPLICANT: Jorgensen, Per Lina  
; APPLICANT: Schulein, Martin  
; TITLE OF INVENTION: FAMILY 44 XYLOGLUCANASES  
; FILE REFERENCE: 10017.200-US  
; CURRENT APPLICATION NUMBER: US/09/784,554B  
; CURRENT FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 4059  
; TYPE: DNA  
; ORGANISM: Paenibacillus polymyxa  
US-09-784-554B-1

## Alignment Scores:

Pred. No.: 6,2e-11 Length: 4059  
Score: 155.50 Matches: 27  
Percent Similarity: 56.47% Conservative: 21  
Best Local Similarity: 31.76% Mismatches: 36  
Query Match: 33.09% Indels: 1  
DB: 4 Gaps: 1

US-09-917-376-4 (1-89) x US-09-784-554B-1 (1-4059)

Qy 4 GlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23  
Db 3598 GGACTGCTGCTCCAATATCGCACAGCAGATCTAAGGTGAACGATAATCACCTCAATCCG 3657  
Qy 24 GlyLeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrValThrValArg 43  
Db 3658 CAATTCCAAATTTGTAACAAAGGCACAACTCCATACCGATCAACGAGTTGAAAATTCGC 3717  
Qy 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63  
Db 3718 TACTACTACACANTCGACGGTGACCGTGAGCAG---ACTTTCACTGCAGCTATGCGAGC 3774  
Qy 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83  
Db 3775 CTGACTGCTCAAGCTCAACGGTAACTGGTTAAATGAGAAGGCTCAACCGGTGCC 3834  
Qy 84 AspThrTyrLeuGln 88  
Db 3835 GATTATTATTGGAA 3849

## RESULT 12

US-09-784-554B-3  
; Sequence 3, Application US/09784554B  
; Patent No. 6815192  
; GENERAL INFORMATION:  
; APPLICANT: Schnorr, Kirk  
; APPLICANT: Jorgensen, Per Lina  
; APPLICANT: Schulein, Martin  
; TITLE OF INVENTION: FAMILY 44 XYLOGLUCANASES  
; FILE REFERENCE: 10017.200-US  
; CURRENT APPLICATION NUMBER: US/09/784,554B  
; CURRENT FILING DATE: 2001-02-16



```
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 4056
; TYPE: DNA
; ORGANISM: Paenibacillus polymyxa
US-09-784-554B-3

Alignment Scores:
Pred. No.: 3 06e-10 Length: 4056
Score: 150.50 Matches: 27
Percent Similarity: 56.4% Conservative: 21
Best Local Similarity: 31.76% Mismatches: 36
Query Match: 32.02% Indels: 1
DB: 4 Gaps: 1

US-09-917-376-4 (1-89) x US-09-784-554B-3 (1-4056)
QY 4 GlyValIysValGlnTyrIleAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
DB 3595 GGATTGGTGCTCCGATATCGCAGCGGATGACAAATGTGAAGCAATCACTTGAACCG 3654
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
DB 3655 CATTTCGAAATTTAAATAAGGTACATCTCGTACCGATCAACGAGTGTGAAAATTGCG 3714
QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63
DB 3715 TACTACTACAGCATCGCGTGACCGTGACGAGCAG---ACATTCACTGCGACTATCGGTG 3771
QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
DB 3772 CTGACTCTCTCGAAGCTGAATGTTAAGTGGTAAATGGATAAAGCTGCAACCGGTGCT 3831
QY 84 AspThrTyrLeuGln 88
DB 3832 GATTATTATTGGAA 3846

RESULT 13
US-07-862-588B-6
; Sequence 6, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linnaa
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 5916796 No. 5916796disk of No. 5916796th America, Inc.
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728

; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1624 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Bacillus lautus
; STRAIN: NCIMB 40250
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 30..1607
; OTHER INFORMATION:
US-07-862-588B-6

Alignment Scores:
Pred. No.: 1.18e-06 Length: 1624
Score: 120.50 Matches: 28
Percent Similarity: 54.12% Conservative: 18
Best Local Similarity: 32.94% Mismatches: 36
Query Match: 25.64% Indels: 3
DB: 2 Gaps: 2

US-09-917-376-4 (1-89) x US-07-862-588B-6 (1-1624)
QY 2 SerGlyGlyValIysValGlnTyrIleAsnAspSerAlaProGlyAspAsnGlnIle 21
DB 1323 ACGGGGAACCTTGTTCATACAAAGTTGGGACACACTAGCCGCGGATACCAATG 1382
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
DB 1383 AGCCCTTCCTTTAATCAATCAAGACACGCTGACACCCCTGTTAACTGAGCGCTCAAG 1442
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
DB 1443 CTTNNNNNNNNNNNNNNAAAAAGAC---GGACCTGCGGATATGAGCTGCTCGACTGG 1499
QY 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81
DB 1500 GCGCAATCGCGCGAACAAGTGTCTGTGGCATTC-----GCTAACTTACGGGAGT 1553
QY 82 ThrAlaAspThrTyr 86
DB 1554 AATACGATACCTTAC 1568

RESULT 14
US-07-862-588B-5
; Sequence 5, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linnaa
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 5916796 No. 5916796disk of No. 5916796th America, Inc.
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
```



;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/07/862,588B

;; FILING DATE: 19920727

;; CLASSIFICATION: 435

;; PRIORITY APPLICATION DATA:  
;; APPLICATION NUMBER: DK 164/90

;; FILING DATE: 19-JAN-1990

;; PRIORITY APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/DK91/00013

;; FILING DATE: 18-JAN-1991

;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Zelson, Steve T. / Lambiris, Elias J.

;; REGISTRATION NUMBER: 30,335 / 33,728

;; REFERENCE/DOCKET NUMBER: 3425.204-US

;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 212 867 0123

;; TELEFAX: 212 867 0298

;; INFORMATION FOR SEQ ID NO: 5:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 1775 base pairs

;; TYPE: NUCLEIC ACID

;; STRANDEDNESS: double

;; TOPOLOGY: linear

;; MOLECULE TYPE: DNA (genomic)

;; FRAGMENT TYPE: N-terminal

;; ORIGINAL SOURCE:

;; ORGANISM: Bacillus lautus

;; STRAIN: NCIMB 40250

;; FEATURE:

;; NAME/KEY: CDS

;; LOCATION: 30..(1625,1775)

;; IDENTIFICATION METHOD: experimental

;; OTHER INFORMATION: /partial

;; OTHER INFORMATION: /evidence= EXPERIMENTAL

;; OTHER INFORMATION: /transl\_except= (pos: 1446 .. 1458, aa:

;; OTHER INFORMATION: OTR)

US-07-862-588B-5

Alignment Scores:  
Pred. No.: 1.34e-06 Length: 1775  
Score: 120.50 Matches: 28  
Percent Similarity: 54.12% Conservative: 18  
Best Local Similarity: 32.94% Mismatches: 36  
Query Match: 25.64% Indels: 3  
DB: 2 Gaps: 2

US-09-917-376-4 (1-89) x US-07-862-588B-5 (1-1775)

QY 2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21  
Db 1323 ACGGGAAACCTGTGTTCCATACAAAGTTGGGACACTAGCGGCACGGATACCAATG 1382  
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41  
Db 1383 AAGCTTCTCTTAACATCAAGAACCAACGGTACACCCCTGTTAACTCGAGCGCTCAAG 1442  
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTirp 61  
Db 1443 CTTNNNNNNNNNNNNAAAAAGAC---GGACCTCGGATATGAGTGTCTCGACTG 1499  
QY 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81  
Db 1500 GCGAAATCGGCGAAGCAATGTTCTGTGTCGATTC-----GCTAACTTTACCGGGAGT 1553  
QY 82 ThrAlaAspThrTyr 86  
Db 1554 AATACGGATACTTAC 1568

RESULT 15

US-08-048-164A-1

;; Sequence 1, App6934

;; Patent No. 5496934

;; GENERAL INFORMATION:

;; APPLICANT: Shoseyov, Oded  
;; APPLICANT: Shpiegl, Itai  
;; APPLICANT: Goldstein, Marc A.  
;; APPLICANT: Doi, Roy H.  
;; TITLE OF INVENTION: CELLULOSE BINDING DOMAIN  
;; NUMBER OF SEQUENCES: 21  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: PENNIE & EDMONDS  
;; STREET: 1155 Avenue of the Americas  
;; CITY: New York  
;; STATE: New York  
;; COUNTRY: U.S.A.  
;; ZIP: 10036  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/048,164A  
;; FILING DATE: 14-APR-1993  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Mistrock, S. Leslie  
;; REGISTRATION NUMBER: 18,872  
;; REFERENCE/DOCKET NUMBER: 7809-003  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (212) 790-9090  
;; TELEFAX: (212) 869-8864/9741  
;; TELEX: 66141 PENNIE  
;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 486 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: unknown  
;; MOLECULE TYPE: DNA  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 1..486  
US-08-048-164A-1

Alignment Scores:  
Pred. No.: 1.65e-06 Length: 486  
Score: 114.00 Matches: 29  
Percent Similarity: 52.58% Conservative: 22  
Best Local Similarity: 29.90% Mismatches: 34  
Query Match: 24.26% Indels: 12  
DB: 1 Gaps: 4

US-09-917-376-4 (1-89) x US-08-048-164A-1 (1-486)

QY 2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21  
Db 7 ACATCATCAATGTCGATTGGAATTTTCAACATCTTAACAAATCAGCACAAACAACTCAATT 66  
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41  
Db 67 ACACCAATTAATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAA 126  
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTirp 61  
Db 127 GTTAGATATTATTACACAAAGTGATGTACACAGGACAAACTTCTGG---TGTGACCAT 183  
QY 62 AlaAlaMetGlyCysGlyAsn-----IleArgAlaSerPhe 73  
Db 184 GCTGTGTCATTATTAGGAAATAGCTATTGTTGATAACACTAGCAAGTGACAGCAAACTTC 243  
QY 74 -----GlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88  
Db 244 GTTAAAGAAACAGCAAGCCCA---ACATCAACCTATGATACATATGTTGAA 291



Search completed: October 5, 2005, 22:56:54  
Job time : 49.3629 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 5, 2005, 08:30:02 ; Search time 171.446 Seconds  
(without alignments)  
3584.301 Million cell updates/sec

Title: US-09-917-376-4  
Perfect score: 470  
Sequence: 1 VSGGVKQYKKNDSAPGDNQ.....RASFGSYNPATPTADTYLQX 89

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 7442561 seqs, 3452328358 residues  
Total number of hits satisfying chosen parameters: 14885122

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters: -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US0917376/runat\_04102005\_164329\_26954/app.query.fasta\_1.2524  
-DB=Published Applications NA -QWTF=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1  
-TRANSL=human40.cdi -LIST=45 -DOCALIGN=200 -ENR=ext -1 -MATRIX=blosum62  
-MAXLEN=200000000 -USR=US0917376 @CGN 1.1 2178 @runat\_04102005\_164329\_26954  
-NCPU=6 -ICPU=3 -NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOC=100  
-LANGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

- 1: /cgn2\_6/ptodata/1/pubna/US07\_PUBCOMB.seq:
- 2: /cgn2\_6/ptodata/1/pubna/PCT\_NEW\_PUB.seq:
- 3: /cgn2\_6/ptodata/1/pubna/US06\_NEW\_PUB.seq:
- 4: /cgn2\_6/ptodata/1/pubna/US06\_PUBCOMB.seq:
- 5: /cgn2\_6/ptodata/1/pubna/US07\_NEW\_PUB.seq:
- 6: /cgn2\_6/ptodata/1/pubna/PCTUS\_PUBCOMB.seq:
- 7: /cgn2\_6/ptodata/1/pubna/US08\_NEW\_PUB.seq:
- 8: /cgn2\_6/ptodata/1/pubna/US08\_PUBCOMB.seq:
- 9: /cgn2\_6/ptodata/1/pubna/US09A\_PUBCOMB.seq:
- 10: /cgn2\_6/ptodata/1/pubna/US09B\_PUBCOMB.seq:
- 11: /cgn2\_6/ptodata/1/pubna/US09C\_PUBCOMB.seq:
- 12: /cgn2\_6/ptodata/1/pubna/US09\_NEW\_PUB.seq:
- 13: /cgn2\_6/ptodata/1/pubna/US10A\_PUBCOMB.seq:
- 14: /cgn2\_6/ptodata/1/pubna/US10B\_PUBCOMB.seq:
- 15: /cgn2\_6/ptodata/1/pubna/US10C\_PUBCOMB.seq:
- 16: /cgn2\_6/ptodata/1/pubna/US10D\_PUBCOMB.seq:
- 17: /cgn2\_6/ptodata/1/pubna/US10E\_PUBCOMB.seq:
- 18: /cgn2\_6/ptodata/1/pubna/US10F\_PUBCOMB.seq:
- 19: /cgn2\_6/ptodata/1/pubna/US10G\_PUBCOMB.seq:
- 20: /cgn2\_6/ptodata/1/pubna/US10H\_PUBCOMB.seq:
- 21: /cgn2\_6/ptodata/1/pubna/US10I\_PUBCOMB.seq:
- 22: /cgn2\_6/ptodata/1/pubna/US10\_NEW\_PUB.seq:
- 23: /cgn2\_6/ptodata/1/pubna/US11A\_PUBCOMB.seq:
- 24: /cgn2\_6/ptodata/1/pubna/US11\_NEW\_PUB.seq:
- 25: /cgn2\_6/ptodata/1/pubna/US60\_NEW\_PUB.seq:
- 26: /cgn2\_6/ptodata/1/pubna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	469	99.8	2289	10 US-09-917-378-2	Sequence 2, Appli
2	466	99.1	3687	10 US-09-917-384-2	Sequence 2, Appli
3	466	99.1	3687	10 US-09-917-383-2	Sequence 2, Appli
4	462	98.3	2869	11 US-09-917-376-2	Sequence 2, Appli
5	462	98.3	2869	15 US-10-155-400-2	Sequence 2, Appli
6	257	54.7	2223	15 US-10-156-761-550	Sequence 550, App
7	257	54.7	9025608	15 US-10-156-761-1	Sequence 11, Appli
8	190	40.4	1527	17 US-10-369-493-46838	Sequence 46838, A
9	188	40.0	1434	19 US-10-466-208-11	Sequence 11, Appli
10	188	40.0	1491	19 US-10-466-208-7	Sequence 7, Appli
11	188	40.0	2510	19 US-10-433-577-34	Sequence 34, Appli
12	165	35.1	501	21 US-10-933-404-1	Sequence 1, Appli
13	165	35.1	1438	17 US-10-372-054-3	Sequence 3, Appli
14	165	35.1	1482	13 US-10-072-152-11	Sequence 11, Appli
15	165	35.1	1482	18 US-10-655-433-11	Sequence 11, Appli
16	157	33.4	1314	9 US-09-974-300-684	Sequence 684, App
17	155.5	33.1	4059	10 US-09-784-554B-1	Sequence 1, Appli
18	155.5	33.1	4059	20 US-10-896-555-1	Sequence 1, Appli
19	150.5	32.0	4056	10 US-09-784-554B-3	Sequence 3, Appli
20	150.5	32.0	4056	20 US-10-896-555-3	Sequence 3, Appli
21	131.5	28.0	3180	22 US-10-886-393A-29	Sequence 29, Appli
22	130.5	27.8	4452	17 US-10-282-122A-15299	Sequence 15299, A
23	126.5	26.9	1303	21 US-10-490-737-1	Sequence 1, Appli
24	126.5	26.9	1747	21 US-10-490-737-2	Sequence 2, Appli
25	126.5	26.9	3489	21 US-10-490-737-3	Sequence 3, Appli
26	114	24.3	768	16 US-10-261-446-5	Sequence 5, Appli
27	114	24.3	768	18 US-10-261-445B-5	Sequence 5, Appli
28	113	24.0	768	21 US-10-782-234-5	Sequence 9, Appli
29	113	24.0	652	20 US-10-748-055-9	Sequence 21, Appli
30	113	24.0	1146	10 US-09-808-898-21	Sequence 21, Appli
31	78.5	16.7	764	20 US-10-653-047-5404	Sequence 5404, Ap
32	78.5	16.7	3147	17 US-10-284-122A-13716	Sequence 13716, A
33	76	16.2	757	19 US-10-767-701-8547	Sequence 8547, Ap
34	76	16.2	96597	18 US-10-052-482-226	Sequence 226, App
35	75.5	16.1	1041	17 US-10-260-238-742	Sequence 742, App
36	75.5	16.1	1041	19 US-10-437-963-76751	Sequence 76751, A
37	75.5	16.1	2091	18 US-10-211-462-222	Sequence 222, App
38	75.5	16.1	2091	20 US-10-723-860-4038	Sequence 4038, Ap
39	75.5	16.1	2091	22 US-10-756-149-4006	Sequence 4006, Ap
40	75.5	16.1	3396	17 US-10-398-458-1	Sequence 1, Appli
41	75.5	16.1	4041	9 US-09-747-835A-21	Sequence 21, Appli
42	75.5	16.1	4041	18 US-10-312-312-21	Sequence 21, Appli
43	75.5	16.1	4353	15 US-10-225-567A-405	Sequence 405, App
44	75.5	16.1	4755	21 US-10-505-486-201	Sequence 201, App
45	75.5	16.1	4914	9 US-09-974-298-188	Sequence 188, App

ALIGNMENTS

RESULT 1  
US-09-917-378-2  
; Sequence 2, Application US/09917378  
; Publication No. US20030119093A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT MANNANASE FROM ACIDOTHERMUS  
; FILE REFERENCE: 40197.7US01  
; CURRENT APPLICATION NUMBER: US/09/917,378  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 8



```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2289
; TYPE: DNA
; ORGANISM: Acidothermus cellulolyticus
US-09-917-378-2

Alignment Scores:
Pred. No.: 5,27e-56 Length: 2289
Score: 469.00 Matches: 88
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.79% Indels: 0
DB: 10 Gaps: 0

US-09-917-376-4 (1-89) x US-09-917-378-2 (1-2289)

Qy 1 ValSerGlyGlyVallyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 1363 GTGTCGGGTGGGTGAGGTGCAGTACAAACAATGATTCGGCGCGGTGATAACCGAG 1422
Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1423 ATCAAAACCGGGTCTCCAGTTGGTGAATACCGGGTCTGTCGGTGGATTTCGACGGTG 1482
Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1483 ACGGTGCGGTACTGGTTCACCGGGATGGTGGGTGCTGCACACTGGTGTACAACCTGTGAC 1542
Qy 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 1543 TGGCGCGCGATGGGTGTGGGAATATCCGCCCTCGTTCGGTCCGTGGAACCCGGCGACG 1602

Qy 81 ProThrAlaAspThrTyrLeuGln 88
Db 1603 CCGACGGCGGACACCTACCTGCAG 1626

RESULT 2
US-09-917-384-2
; Sequence 2, Application US/09917384
; Publication No. US20030096342A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,384
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 3687
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-384-2

Alignment Scores:
Pred. No.: 2,66e-55 Length: 3687
Score: 466.00 Matches: 87
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 98.86% Mismatches: 0
Query Match: 99.15% Indels: 0
DB: 10 Gaps: 0

US-09-917-376-4 (1-89) x US-09-917-384-2 (1-3687)

Qy 1 ValSerGlyGlyVallyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 1750 GTGTCGGGTGGGTGAGGTGCAGTACAAACAATGATTCGGCGCGGTGATAACCGAG 1809
Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1810 ATCAAAACCGGGTCTCCAGTTGGTGAATACCGGGTCTGTCGGTGGATTTCGACGGTG 1869
Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1870 ACGGTGCGGTACTGGTTCACCGGGATGGTGGGTGCTGCACACTGGTGTACAACCTGTGAC 1929
Qy 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 1930 TGGCGCGCGATGGGTGTGGGAATATCCGCCCTCGTTCGGTCCGTGGAACCCGGCGACG 1989
Qy 81 ProThrAlaAspThrTyrLeuGln 88

US-09-917-376-4 (1-89) x US-09-917-383-2 (1-3687)

Qy 1 ValSerGlyGlyVallyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 1750 GTGTCGGGTGGGTGAGGTGCAGTACAAACAATGATTCGGCGCGGTGATAACCGAG 1809
Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1810 ATCAAAACCGGGTCTCCAGTTGGTGAATACCGGGTCTGTCGGTGGATTTCGACGGTG 1869
Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1870 ACGGTGCGGTACTGGTTCACCGGGATGGTGGGTGCTGCACACTGGTGTACAACCTGTGAC 1929
Qy 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 1930 TGGCGCGCGATGGGTGTGGGAATATCCGCCCTCGTTCGGTCCGTGGAACCCGGCGACG 1989
Qy 81 ProThrAlaAspThrTyrLeuGln 88
```



```
Db 1990 CCGACGGCGGACACCTACCTGCAG 2013
|||||
RESULT 4
US-09-917-376-2
; Sequence 2, Application US/09917376
; Publication No. US20040038334A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOYTICUS
; FILE REFERENCE: 40197.4US01
; CURRENT APPLICATION NUMBER: US/09/917,376
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2869
; TYPE: DNA
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (2869)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-10-155-400-2
Alignment Scores:
Pred. No.: 7,09e-55 Length: 2869
Score: 462.00 Matches: 86
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 97.73% Mismatches: 0
Query Match: 98.30% Indels: 0
DB: 15 Gaps: 0
US-09-917-376-4 (1-89) x US-10-155-400-2 (1-2869)
QY 1 ValSerGlyValIysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 2605 GTGTCGGGTGGGTGAGGTGCAGTATAGATATGATTCGGCGCGGGTGATATCAG 2664
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 2665 ATCAAGCCGGTTTCAGGTGGTGAATACCGGGTCGTCGCGTGGATTTGTGCGCGGTG 2724
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 2725 ACGGTGCGGTACTGGTTTCACCGGGATGGTGGCTCGTCGACACTGGTGTAACCTGTGAC 2784
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 2785 TGGCGCGCGCATCGGGTGGGAAATATCCCGCCTCGTTCCGCTCGGTGAACCCGCGCAGC 2844
QY 81 ProThrAlaAspThrTyrLeuGln 88
Db 2845 CCGACGGCGGACACCTACCTGCAG 2868
RESULT 6
US-10-156-761-550
; Sequence 550, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 550
; LENGTH: 2223
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1) .. (2223)
US-10-156-761-550
Alignment Scores:
```







```
; Sequence 11, Application US/10466208
; Publication No. US20040180348A1
; GENERAL INFORMATION:
; APPLICANT: GENOFOCUS Co., Ltd.
; TITLE OF INVENTION: Method for Surface Display of Proteins on Genetic Carriers
; FILE REFERENCE: 3260-13
; CURRENT APPLICATION NUMBER: US/10/466,208
; CURRENT FILING DATE: 2003-07-14
; PRIOR APPLICATION NUMBER: KR2001-2156
; PRIOR FILING DATE: 2001-01-15
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Kopatentin 1.71
; SEQ ID NO 11
; LENGTH: 1434
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: CMCCase gene with additional his encoding sequences
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1431)
US-10-466-208-11
Alignment Scores:
Pred. No.: 3,44e-16 Length: 1434
Score: 188.00 Matches: 34
Percent Similarity: 61.18% Conservatives: 18
Best Local Similarity: 40.00% Mismatches: 31
Query Match: 40.00% Indels: 2
DB: 19 Gaps: 1
US-09-917-376-4 (1-89) x US-10-466-208-11 (1-1434)
Qy 4 GlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 991 GGTATTCTGTACAAATACAGACGAGGGGATGGAGTATGAACAGCAACCAATCGTCCG 1050
Qy 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 1051 CAGCTTCAATATAAATAACGCAATACCAAGGTTGATTAAAGATGTCTACTGCCCGT 1110
Qy 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63
Db 1111 TACTGGTATAACGCAAAACAAAGCCAAACGTT-----GACTGTGACTACGCGCAG 1164
Qy 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 1165 CTTGGATGCGCAATGTGCATACACAAGTTTGTGACGTTGTCATTAACCAAGCAAGGTGCA 1224
Qy 84 AspThrTyrLeuGln 88
Db 1225 GATACCTATCTGGAA 1239
RESULT 10
US-10-466-208-7
; Sequence 7, Application US/10466208
; Publication No. US20040180348A1
; GENERAL INFORMATION:
; APPLICANT: GENOFOCUS Co., Ltd.
; TITLE OF INVENTION: Method for Surface Display of Proteins on Genetic Carriers
; FILE REFERENCE: 3260-13
; CURRENT APPLICATION NUMBER: US/10/466,208
; CURRENT FILING DATE: 2003-07-14
; PRIOR APPLICATION NUMBER: KR2001-2156
; PRIOR FILING DATE: 2001-01-15
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Kopatentin 1.71
; SEQ ID NO 7
; LENGTH: 1491
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: CMCCase gene with mutated signal sequence to enhance its
; OTHER INFORMATION: hydrophobicity
; NAME/KEY: CDS
; LOCATION: (1)..(1488)
US-10-466-208-7
Alignment Scores:
Pred. No.: 3,62e-16 Length: 1491
Score: 188.00 Matches: 34
Percent Similarity: 61.18% Conservatives: 18
Best Local Similarity: 40.00% Mismatches: 31
Query Match: 40.00% Indels: 2
DB: 19 Gaps: 1
US-09-917-376-4 (1-89) x US-10-466-208-7 (1-1491)
Qy 4 GlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 1048 GGTATTCTGTACAAATACAGACGAGGGGATGGAGTATGAACAGCAACCAATCGTCCG 1107
Qy 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 1108 CAGCTTCAATATAAATAACGCAATACCAAGGTTGATTAAAGATGTCTACTGCCCGT 1167
Qy 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63
Db 1168 TACTGGTATAACGCAAAACAAAGCCAAACGTT-----GACTGTGACTACGCGCAG 1221
Qy 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 1222 CTTGGATGCGCAATGTGCATACACAAGTTTGTGACGTTGTCATTAACCAAGCAAGGTGCA 1281
Qy 84 AspThrTyrLeuGln 88
Db 1282 GATACCTATCTGGAA 1296
RESULT 11
US-10-433-577-34
; Sequence 34, Application US/10433577
; Publication No. US20040171065A1
; GENERAL INFORMATION:
; APPLICANT: Genofocus Co., Ltd.
; TITLE OF INVENTION: Method for Expression of Proteins on Spore Surface
; FILE REFERENCE: PCT-Genofocus-1
; CURRENT APPLICATION NUMBER: US/10/433,577
; CURRENT FILING DATE: 2003-06-05
; PRIOR APPLICATION NUMBER: KR2000-74835
; PRIOR FILING DATE: 2000-12-08
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: Kopatentin 1.71
; SEQ ID NO 34
; LENGTH: 2510
; TYPE: DNA
; ORGANISM: Bacillus subtilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (461)..(2491)
US-10-433-577-34
Alignment Scores:
Pred. No.: 7,25e-16 Length: 2510
Score: 188.00 Matches: 34
Percent Similarity: 61.18% Conservatives: 18
Best Local Similarity: 40.00% Mismatches: 31
Query Match: 40.00% Indels: 2
DB: 19 Gaps: 1
US-09-917-376-4 (1-89) x US-10-433-577-34 (1-2510)
Qy 4 GlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 2051 GGTATTCTGTACAAATACAGACGAGGGGATGGAGTATGAACAGCAACCAATCGTCCG 2110
```



```

Qy 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 2111 CAGCTTCAATAAATAACGGCAATACACAGCTTGATTTAAAGATGTCACCTGCCGT 2170
Qy 44 TyrTrpPheThrArgAspGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63
Db 2171 TACTGGTATACCGCAAAACAAAGCGCAAAACGTT-----GACTGTGACTACGCGCAG 2224
Qy 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 2225 CTTGGATCGCGCAATGTCATACATACAGTTTGTGACGTTTGCAATAAACCAAGCAAGGTGCA 2284
Qy 84 AspThrTyrLeuGln 88
Db 2285 GATACCTATCTGGAA 2299

RESULT 12
US-10-933-404-1
; Sequence 1, Application US/10933404
; Publication No. US20050118729A1
; GENERAL INFORMATION:
; APPLICANT: MORAG, Ely
; TITLE OF INVENTION: MICROARRAYS OF CELLULOSE BINDING CHIMERIC PROTEINS AND METHODS OF
; FILE OF INVENTION: THEREOF
; FILE REFERENCE: 85189-7100
; CURRENT APPLICATION NUMBER: US/10/933,404
; CURRENT FILING DATE: 2004-09-03
; PRIOR APPLICATION NUMBER: PCT/IL03/00177
; PRIOR FILING DATE: 2003-03-06
; PRIOR APPLICATION NUMBER: US 60/362,061
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Clostridium thermocellum
US-10-933-404-1

Alignment Scores:
Pred. No.: 1.62e-13 Length: 501
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.11% Indels: 10
DB: 21

US-09-917-376-4 (1-89) x US-10-933-404-1 (1-501)

Qy 1 ValSerGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 13 GTATCAGGCAATTTGAAGTTGAATCTTACACAGCAATCTCTCAGATACCTAACTCA 72
Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 73 ATCAATCTCAGTTCAAGTTACTAATACCGAAGCAGTGAATGATTGTCCAAACTC 132
Qy 41 ThrValArgTyrTrpPheThrArgAspGlySerSerThrLeuValTyrAsnCysAsp 60
Db 133 ACATTGAGATATTATTATACAGTAGACGACGAGAAAGATCAGACCTCTCGG---TGTGAC 189
Qy 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 190 CATGCTGCAATAATCGGCAAGTACACGAGCAATCTTCAATGTAAAGGA 249
Qy 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 250 ACATTGTAAATAAGTTCCTCAACAAATAACGACGACACCTACCTTGAA 300

RESULT 13
US-10-372-054-3
; Sequence 3, Application US/10372054
; Publication No. US20030203466A1

```

```

; GENERAL INFORMATION:
; APPLICANT: Kauppinen, Markus
; APPLICANT: Schulein, Martin
; APPLICANT: Schnorr, Kirk
; APPLICANT: Andersen, Lene
; APPLICANT: Bjornvad, Mads
; TITLE OF INVENTION: Novel Mannanases
; FILE REFERENCE: 5440.204-US
; CURRENT APPLICATION NUMBER: US/10/372,054
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: US/09/339,159B
; PRIOR FILING DATE: 1999-06-24
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1438
; TYPE: DNA
; ORGANISM: Bacillus sp. i633
US-10-372-054-3

Alignment Scores:
Pred. No.: 6.59e-13 Length: 1438
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.11% Indels: 10
DB: 17 2

US-09-917-376-4 (1-89) x US-10-372-054-3 (1-1438)

Qy 1 ValSerGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 940 GTATCAGGCAATTTGAAGTTGAATCTTACACAGCAATCTCTCAGATACCTAACTCA 999
Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1000 ATCAATCTCAGTTCAAGTTACTAATACCGAAGCAGTGAATGATTGTCCAAACTC 1059
Qy 41 ThrValArgTyrTrpPheThrArgAspGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1060 ACATTGAGATATTATTATACAGTAGACGACGACGAGAAAGATCAGACCTCTCGG---TGTGAC 1116
Qy 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1117 CATGCTGCAATAATCGGCAAGTACACGAGCAATCTTCAATGTAAAGGA 1176
Qy 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1177 ACATTGTAAATAAGTTCCTCAACAAATAACGACGACACCTACCTTGAA 1227

RESULT 14
US-10-072-152-11
; Sequence 11, Application US/10072152
; Publication No. US20020142438A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. US20020142438A1el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/10/072,152
; CURRENT FILING DATE: 2002-02-07
; PRIOR APPLICATION NUMBER: US/09/198,955
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97

```



```
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Clostridium thermocellum
US-10-655-433-11

Alignment Scores:
Pred. No.: 6,86e-13 Length: 1482
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.11% Indels: 10
DB: 18 Gaps: 2

US-09-917-376-4 (1-89) x US-10-655-433-11 (1-1482)
QY 1 ValSerGlyValValValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGCAATTTCAGGTTGAATCTACACAGCAATCTTCAGATACTACTAACTCA 1050
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCCTCAGTTCAGGTTACTAATACCGAAGCAGTGAATTTGTTCCAAACTC 1110
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATTATACAGTAGACGACAGAAAGATCAGACCTTCTGG---TGTGAC 1167
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCAATATTCGGCAGTAGTAACGGCAGCTACAACGGAATTACTTCAATGTAAAGGA 1227
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTGTGTAATAATGAGTTCTCTCAACAAATAACGCAGACACCTACCTTGAA 1278

Search completed: October 6, 2005, 04:57:57
Job time : 911.446 secs
```

```
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Clostridium thermocellum
US-10-072-152-11

Alignment Scores:
Pred. No.: 6,86e-13 Length: 1482
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.11% Indels: 10
DB: 13 Gaps: 2

US-09-917-376-4 (1-89) x US-10-072-152-11 (1-1482)
QY 1 ValSerGlyValValValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGCAATTTCAGGTTGAATCTACACAGCAATCTTCAGATACTACTAACTCA 1050
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCCTCAGTTCAGGTTACTAATACCGAAGCAGTGAATTTGTTCCAAACTC 1110
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATTATACAGTAGACGACAGAAAGATCAGACCTTCTGG---TGTGAC 1167
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCAATATTCGGCAGTAGTAACGGCAGCTACAACGGAATTACTTCAATGTAAAGGA 1227
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTGTGTAATAATGAGTTCTCTCAACAAATAACGCAGACACCTACCTTGAA 1278
```

```
RESULT 15
US-10-655-433-11
; Sequence 11, Application US/10655433
; Publication No. US20040067572A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. US20040067572A1el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/10/655,433
; PRIOR APPLICATION NUMBER: US/09/198,955A
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 07:48:48 ; Search time 16 8581 Seconds  
(without alignments)  
2018.911 Million cell updates/sec

Title: US-09-917-376-5

Perfect score: 459

Sequence: 1 VSGGVKQVKNDSAPGDNQ.....IRASFGSVNPATPTADTYLQ 88

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_16Dec04.\*

1: geneseqp1980s.\*

2: geneseqp1990s.\*

3: geneseqp2000s.\*

4: geneseqp2001s.\*

5: geneseqp2002s.\*

6: geneseqp2003as.\*

7: geneseqp2003bs.\*

8: geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	469	100.0	88	7 ADD22925	Add22925 Acidother
2	469	100.0	89	7 ADD22924	Add22924 Acidother
3	469	100.0	154	6 ABP73020	Abp73020 Amino aci
4	469	100.0	154	7 ADJ38294	Adj38294 A. cellul
5	469	100.0	762	6 ABP73022	Abp73022 Amino aci
6	469	100.0	762	7 ADJ38291	Adj38291 A. cellul
7	466	99.4	150	6 ABP73025	Abp73025 Amino aci
8	466	99.4	150	8 ADH36640	Adh36640 A. cellul
9	466	99.4	1228	6 ABP73029	Abp73029 Amino aci
10	466	99.4	1228	8 ADH36636	Adh36636 Acidother
11	462	98.5	88	6 ABP73018	Abp73018 Amino aci
12	462	98.5	88	6 AD52317	Ad52317 A. cellul
13	462	98.5	89	6 ABP73017	Abp73017 Amino aci
14	462	98.5	89	8 AD52316	Ad52316 A. cellul
15	462	98.5	957	6 ABP73015	Abp73015 Amino aci
16	462	98.5	957	7 ADD22921	Add22921 Acidother
17	462	98.5	957	8 AD52313	Ad52313 Acidother
18	436	93.0	153	6 ABP71658	Abp71658 A. cellul
19	436	93.0	153	7 ADP75898	Adp75898 Carbohyd
20	436	93.0	1121	6 ABP71656	Abp71656 A. cellul
21	436	93.0	1121	7 ADP75895	Adp75895 Acidother
22	231.5	49.4	782	2 AAR15625	Aar15625 Cellulase
23	195.5	41.7	616	2 AAY13494	Aay13494 Truncated
24	195.5	41.7	616	5 AAE16325	Aae16325 Active ce
25	194.5	41.5	1751	2 AAY13493	Aay13493 Truncated

#### ALIGNMENTS

##### RESULT 1

ADD22925

ID ADD22925 standard; protein; 88 AA.

XX

AC ADD22925;

XX

DT 15-JAN-2004 (first entry)

XX

DE Acidothermus cellulolyticus avicelase AvIII CBD III #2.

XX

KW enzyme; AvIII; cellulose reduction; agricultural biomass;

KW municipal solid waste; glycoside hydrolase; avicelase.

XX

OS Acidothermus cellulolyticus.

XX

PN US2003108988-A1.

XX

PD 12-JUN-2003.

XX

PF 18-OCT-2002; 2002US-00155400.

XX

PR 28-JUL-2001; 2001US-00917376.

XX

PA (DING/) DING S.

PA (ADNE/) ADNEY W S.

PA (VINZ/) VINZANT T B.

PA (HIMM/) HIMMEL M B.

PI

Ding S, Adney WS, Vinzant TB, Himmel MB;

XX

DR WPI; 2003-810853/76.

XX

New isolated thermal tolerant avicelase polynucleotide useful for detection of a polynucleotide encoding AvIII and for reducing cellulose in a starting material, e.g. municipal solid waste.

PT

Claim 16; SEQ ID NO 5; 29pp; English.

PS

The invention relates to an isolated polynucleotide molecule encoding a thermostable AvIII polypeptide. The polynucleotide is useful for detection of a polynucleotide encoding AvIII. The polynucleotide is useful for reducing cellulose in a starting material which involves administering to the starting material, e.g. agricultural biomass or municipal solid waste, a polypeptide molecule of the polynucleotide. The method further comprises administering a second polypeptide molecule chosen from the glycoside hydrolase family of proteins. The present sequence represents the amino acid sequence of Acidothermus cellulolyticus avicelase AvIII CBD III #2.

CC



XX SQ Sequence 88 AA;

Query Match 100.0%; Score 469; DB 7; Length 88;  
 Best Local Similarity 100.0%; Pred. No. 5.4e-46;  
 Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKVOYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRWFTRDGGSTLVNCD 60  
 DB 1 VSGGVKVOYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRWFTRDGGSTLVNCD 60

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
 DB 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88

RESULT 2  
 ADD22924  
 ID ADD22924 standard; protein; 89 AA.  
 XX  
 AC ADD22924;  
 DT 15-JAN-2004 (first entry)  
 DE Acidothermus cellulolyticus avicelase AvIII CBD III #1.  
 KW enzyme; AvIII; cellulose reduction; agricultural biomass;  
 KW municipal solid waste; glycoside hydrolase; avicelase.  
 XX  
 OS Acidothermus cellulolyticus.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 89 /label= Unknown  
 FT  
 XX  
 PN US2003108988-A1.  
 PD 12-JUN-2003.  
 XX  
 PF 18-OCT-2002; 2002US-00155400.  
 XX  
 PR 28-JUL-2001; 2001US-00917376.  
 XX  
 PA (DING/) DING S.  
 PA (ADNE/) ADNEY W S.  
 PA (VINZ/) VINZANT T B.  
 PA (HIMM/) HIMMEL M E.  
 XX  
 PI Ding S, Adney WS, Vinzant TB, Himmel ME;  
 XX  
 DR WPI; 2003-810853/76.  
 XX  
 PT New isolated thermal tolerant avicelase polynucleotide useful for  
 PT detection of a polynucleotide encoding AvIII and for reducing cellulose  
 PT in a starting material, e.g. municipal solid waste.  
 XX  
 PS Claim 16; SEQ ID NO 4; 29pp; English.  
 XX  
 CC The invention relates to an isolated polynucleotide molecule encoding a  
 CC thermostable AvIII polypeptide. The polynucleotide is useful for  
 CC detection of a polynucleotide encoding AvIII. The polynucleotide is  
 CC useful for reducing cellulose in a starting material which involves  
 CC administering to the starting material, e.g. agricultural biomass or  
 CC municipal solid waste, a polypeptide molecule of the polynucleotide. The  
 CC method further comprises administering a second polypeptide molecule  
 CC chosen from the glycoside hydrolase family of proteins. The present  
 CC sequence represents the amino acid sequence of Acidothermus  
 CC cellulolyticus avicelase AvIII CBD III #1.  
 XX  
 SQ Sequence 89 AA;

Query Match 100.0%; Score 469; DB 7; Length 89;  
 Best Local Similarity 100.0%; Pred. No. 5.5e-46;

Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKVOYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRWFTRDGGSTLVNCD 60  
 DB 1 VSGGVKVOYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRWFTRDGGSTLVNCD 60

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
 DB 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88

RESULT 3  
 ABP73020  
 ID ABP73020 standard; peptide; 154 AA.  
 XX  
 AC ABP73020;  
 DT 03-JUN-2003 (first entry)  
 DE Amino acid sequence of the ManA carbohydrate binding domain type III.  
 KW ManA; mannanase A; glycoside hydrolase; enzyme; hemicellulose; sugar;  
 KW food; feed; paper pulp; biofuel; mannase.  
 XX  
 OS Acidothermus cellulolyticus.  
 XX  
 PN WO2003012110-A1.  
 XX  
 PD 13-FEB-2003.  
 XX  
 PF 28-JUL-2001; 2001WO-US023819.  
 XX  
 PR 28-JUL-2001; 2001WO-US023819.  
 XX  
 PA (MIDE ) MIDWEST RES INST.  
 XX  
 PI Ding S, Adney WS, Vinzant TB, Himmel ME;  
 XX  
 DR WPI; 2003-248182/24.  
 XX  
 PT Novel thermal tolerant mannanase A polypeptide derived from Acidothermus  
 PT cellulolyticus, useful for reducing hemicellulose in a starting material,  
 PT for processing of food, and as bulking agents in food stuffs.  
 XX  
 PS Claim 8; Page 7; 46pp; English.  
 XX  
 CC The present sequence represents a fragment of ManA, a thermostable  
 CC mannanase A polypeptide derived from Acidothermus cellulolyticus. ManA is  
 CC a member of the glycoside hydrolase family of enzymes. ManA is useful for  
 CC reducing hemicellulose in a starting material to simpler carbohydrate  
 CC units, and ultimately to sugars which are useful in the food, feed, paper  
 CC pulp, and biofuels industries. It is useful for the processing of food  
 CC and in food stuffs as bulking agents, and for the degradation of mannan.  
 CC ManA is also useful to raise polyclonal and monoclonal antibodies that  
 CC are useful in purifying ManA, or detecting ManA polypeptide expression,  
 CC and as well as reagent tools for characterizing the molecular actions of  
 CC ManA polypeptides  
 XX  
 SQ Sequence 154 AA;

Query Match 100.0%; Score 469; DB 6; Length 154;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-45;  
 Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKVOYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRWFTRDGGSTLVNCD 60  
 DB 1 VSGGVKVOYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRWFTRDGGSTLVNCD 60

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
 DB 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88







```
XX 28-JUL-2001; 2001US-00917378.
PR (DING/) DING S.
PA (ADNE/) ADNEY W S.
PA (VINZ/) VINZANT T B.
PA (HIMM/) HIMMEL M E.
XX
PI Ding S, Adney WS, Vinzant TB, Himmel ME;
XX
DR WPI; 2003-863435/80.
DR N-ESDB; ADJ38292.
XX
XX New composition comprising a purified mannanase A peptide, comprising a
PT catalytic domain GH5, carbohydrate binding domain III or carbohydrate
PT binding domain II, useful in food processing and in food stuffs as
PT bulking agents.
XX
PS Claim 10; SEQ ID NO 1; 18pp; English.
XX
CC The invention relates to a composition comprising a purified mannanase A
CC peptide, having a catalytic domain GH5, carbohydrate binding domain III
CC or carbohydrate binding domain II. The composition is useful in food
CC processing and in foodstuffs as bulking agents. The present sequence
CC represents the amino acid sequence of A. cellulolyticus mannanase ManA.
XX
XX Sequence 762 AA;
SQ
Query Match 100.0%; Score 469; DB 7; Length 762;
Best Local Similarity 100.0%; Pred. No. 7.8e-45;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 VSGGVKVOYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYVWFTRDGGSSTLVNCD 60
DB 455 VSGGVKVOYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYVWFTRDGGSSTLVNCD 514
OY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
DB 515 WAAMGCGNIRASFGSVNPATPTADTYLQ 542
XX
RESULT 7
ABP73025
ID ABP73025 standard; peptide; 150 AA.
XX
AC ABP73025;
XX
DT 03-JUN-2003 (first entry)
XX
DE Amino acid sequence of the GuxA carbohydrate binding domain type III.
XX
KW GuxA; cellulase; glycoside hydrolase; enzyme; cellulose; sugar; biofuel;
KW detergent; pulp processing; paper processing; feed processing; textile.
XX
OS Acidothermus cellulolyticus.
XX
PN WO2003012109-A1.
XX
PD 13-FEB-2003.
XX
PF 28-JUL-2001; 2001WO-US023817.
XX
PR 28-JUL-2001; 2001WO-US023817.
XX
PA (MIDE ) MIDWEST RES INST.
XX
PI Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;
XX
DR WPI; 2003-239526/23.
XX
XX Novel thermal tolerant GuxA polypeptide derived from Acidothermus
PT cellulolyticus, useful for reducing cellulose in a starting material, and
PT for the conversion of biomass to biofuels and biofuel additives.
XX
XX Claim 6; Page 7; 47pp; English.
XX
CC ABP73023-28 represent peptide fragments of a GuxA polypeptide. GuxA is a
CC thermostable cellulase, and is a member of the glycoside hydrolase family
CC of enzymes. GuxA is useful for reducing cellulose in a starting material
CC such as agricultural biomass to sugars. This is useful in biofuel
CC production. GuxA is also useful in the conversion of biomass to biofuels
CC and biofuel additives, in detergents, pulp and paper processing, food and
CC feed processing, and in textile process. GuxA is also useful for raising
CC polyclonal and monoclonal antibodies that are useful in purifying GuxA,
CC or detecting GuxA polypeptide expression, as well as reagent tools for
CC characterizing the molecular actions of GuxA polypeptides
XX
XX Sequence 150 AA;
SQ
Query Match 99.4%; Score 466; DB 6; Length 150;
Best Local Similarity 98.9%; Pred. No. 2.3e-45;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 1 VSGGVKVOYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYVWFTRDGGSSTLVNCD 60
DB 1 VSGGLKVOYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYVWFTRDGGSSTLVNCD 60
OY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
DB 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
XX
RESULT 8
ADH36640
ID ADH36640 standard; protein; 150 AA.
XX
AC ADH36640;
XX
DT 11-MAR-2004 (first entry)
XX
DE A. cellulolyticus glycoside hydrolase, GuxA CBD type III domain.
XX
KW Thermal tolerant cellulase; glycoside hydrolase; GuxA;
KW carbohydrate degradation; cellulose; agricultural biomass;
KW municipal solid waste; thermostable; enzyme;
KW carbohydrate binding domain type III; CBD III.
XX
OS Acidothermus cellulolyticus.
XX
PN US2003104522-A1.
XX
PD 05-JUN-2003.
XX
PF 28-JUL-2001; 2001US-00917383.
XX
PR 28-JUL-2001; 2001US-00917383.
XX
PA (DING/) DING S.
PA (ADNE/) ADNEY W S.
PA (VINZ/) VINZANT T B.
PA (HIMM/) HIMMEL M E.
PA (DECK/) DECKER S R.
XX
PI Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;
XX
DR WPI; 2004-106451/11.
XX
XX Novel isolated thermostable GuxA polypeptide useful for detecting
PT polynucleotide encoding GuxA, assessing carbohydrate degradation activity
PT of GuxA, reducing cellulose in starting material e.g., agricultural
PT biomass.
XX
XX Claim 27; SEQ ID NO 5; 20pp; English.
XX
XX The present invention relates to the isolation of a thermal tolerant
CC cellulase from Acidothermus cellulolyticus. The cellulase is a member of
```



CC the glycoside hydrolase family and is designated GuxA. Also disclosed are  
 CC methods of producing and using GuxA. The GuxA polypeptide is useful for  
 CC the detection of a polynucleotide encoding GuxA. The polypeptide sequence  
 CC is also useful for assessing the carbohydrate, e.g. cellulose,  
 CC degradation activity of GuxA. GuxA is useful for reducing cellulose in a  
 CC starting material such as agricultural biomass or municipal solid waste.  
 CC The polypeptide molecule of GuxA is thermostable. The present sequence  
 CC represents a domain of the A. cellulolyticus GuxA polypeptide.  
 XX  
 SQ Sequence 150 AA;

Query Match 99.4%; Score 466; DB 8; Length 150;  
 Best Local Similarity 98.9%; Pred. No. 2.3e-45;  
 Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYVWFTRDGGSSTLVYVNC 60  
 DB 1 VSGGLKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYVWFTRDGGSSTLVYVNC 60  
 QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88  
 DB 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88

RESULT 9  
 ABP73029  
 ID ABP73029 standard; protein; 1228 AA.

XX  
 AC ABP73029;

DT 03-JUN-2003 (first entry)

XX Amino acid sequence of the GuxA polypeptide.

XX GuxA; cellulase; glycoside hydrolase; enzyme; cellulose; sugar; biofuel;  
 KW detergent; pulp processing; paper processing; feed processing; textile.

XX Acidothermus cellulolyticus.

XX WO2003012109-A1.

XX 13-FEB-2003.

XX 28-JUL-2001; 2001WO-US023817.

XX 28-JUL-2001; 2001WO-US023817.

XX (MIDE ) MIDWEST RES INST.

XX Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;

XX WPI; 2003-239526/23.

XX N-PSDB; ABZ77634.

XX Novel thermal tolerant GuxA polypeptide derived from Acidothermus  
 PT cellulolyticus, useful for reducing cellulose in a starting material, and  
 PT for the conversion of biomass to biofuels and biofuel additives.

XX Claim 3; Page 19; 47pp; English.

XX The present sequence represents a GuxA polypeptide. GuxA is thermostable  
 CC cellulase, and is a member of the glycoside hydrolase family of enzymes.  
 CC GuxA is useful for reducing cellulose in a starting material such as  
 CC agricultural biomass to sugars. This is useful in biofuel production.  
 CC GuxA is also useful in the conversion of biomass to biofuels and biofuel  
 CC additives, in detergents, pulp and paper processing, food and feed  
 CC processing, and in textile process. GuxA is also useful for raising  
 CC polyclonal and monoclonal antibodies that are useful in purifying GuxA,  
 CC or detecting GuxA polypeptide expression, as well as reagent tools for  
 CC characterizing the molecular actions of GuxA polypeptides

XX Sequence 1228 AA;

Query Match 99.4%; Score 466; DB 6; Length 1228;  
 Best Local Similarity 98.9%; Pred. No. 3.1e-44;  
 Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VSGGVKVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYVWFTRDGGSSTLVYVNC 60  
 DB 584 VSGGLKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYVWFTRDGGSSTLVYVNC 643  
 QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88  
 DB 644 WAAMGCGNIRASFGSVNPATPTADTYLQ 671

RESULT 10  
 ADH36636

ID ADH36636 standard; protein; 1228 AA.

XX  
 AC ADH36636;

XX 11-MAR-2004 (first entry)

XX Acidothermus cellulolyticus glycoside hydrolase, GuxA.

XX Thermal tolerant cellulase; glycoside hydrolase; GuxA;  
 KW carbohydrate degradation; cellulose; agricultural biomass;  
 KW municipal solid waste; thermostable; enzyme.

XX Acidothermus cellulolyticus.

XX US2003104522-A1.

XX 05-JUN-2003.

XX 28-JUL-2001; 2001US-00917383.

XX 28-JUL-2001; 2001US-00917383.

XX (DING/) DING S.

XX (ADNE/) ADNEY W S.

XX (VINZ/) VINZANT T B.

XX (HIMM/) HIMMEL M E.

XX (DECK/) DECKER S R.

XX Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;

XX WPI; 2004-106451/11.

XX N-PSDB; ADH36637.

XX Novel isolated thermostable GuxA polypeptide useful for detecting  
 PT polynucleotide encoding GuxA, assessing carbohydrate degradation activity  
 PT of GuxA, reducing cellulose in starting material e.g., agricultural  
 PT biomass.

XX Claim 27; SEQ ID NO 1; 20pp; English.

XX The present invention relates to the isolation of a thermal tolerant  
 CC cellulase from Acidothermus cellulolyticus. The cellulase is a member of  
 CC the glycoside hydrolase family and is designated GuxA. Also disclosed are  
 CC methods of producing and using GuxA. The GuxA polypeptide is useful for  
 CC the detection of a polynucleotide encoding GuxA. The polypeptide sequence  
 CC is also useful for assessing the carbohydrate, e.g. cellulose,  
 CC degradation activity of GuxA. GuxA is useful for reducing cellulose in a  
 CC starting material such as agricultural biomass or municipal solid waste.  
 CC The polypeptide molecule of GuxA is thermostable. The present sequence  
 CC represents A. cellulolyticus glycoside hydrolase, GuxA.

XX Sequence 1228 AA;

Query Match 99.4%; Score 466; DB 8; Length 1228;  
 Best Local Similarity 98.9%; Pred. No. 3.1e-44;  
 Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYVWFTRDGGSSTLVYVNC 60



Db 584 VSGGLKVOYKNDSSAPGDNQIKPGLQVNTGSSVDLSTVTYRWFTRDGSSTLVNCD 643  
Qy 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
Db 644 WAAMCGNIRASFGSVNPATPTADTYLQ 671

RESULT 11  
ABP73018  
ID ABP73018 standard; peptide; 88 AA.  
AC ABP73018;  
XX  
XX 03-JUN-2003 (first entry)  
DE Amino acid sequence of AvIII carbohydrate binding domain type II.  
XX  
XX Avicelase; AvIII; glycoside hydrolase; enzyme; cellulase; biofuel;  
KW detergent; pulp processing; paper processing; feed processing; textile;  
KW cellulose.  
XX  
XX Acidothermus cellulolyticus.  
OS  
XX WO2003012090-A2.  
XX  
XX 13-FEB-2003.  
XX  
XX 28-JUL-2001; 2001WO-US023818.  
XX  
XX 28-JUL-2001; 2001WO-US023818.  
XX  
XX (MIDE ) MIDWEST RES INST.  
PA  
XX Ding S, Adney WS, Vinzant TB, Himmel ME;  
PI WPI; 2003-248177/24.  
DR  
XX New thermostable AvIII peptide from Acidothermus cellulolyticus, useful  
PT for degradation of cellulose or in generating anti-AvIII antibodies for  
PT purifying recombinant AvIII polypeptides from genetically engineered  
PT host cells.  
XX  
XX Claim 6; Page 8; 44pp; English.  
XX  
XX The present sequence is derived from a thermostable avicelase, designated  
CC AvIII. AvIII is a member of the glycoside hydrolase family of enzymes,  
CC and is a cellulase. AvIII is useful in the conversion of biomass to  
CC biofuels and biofuel additives. It may be useful in the production of  
CC detergents, pulp and paper processing, food and feed processing and in  
CC textile processes. The thermostable AvIII peptide is useful in the  
CC degradation of cellulose, and in generating specific anti-AvIII  
CC antibodies that are useful in purifying recombinant AvIII polypeptides  
CC from genetically engineered host cells, in detecting AvIII polypeptide  
CC expression, as well as a reagent tool for characterizing the molecular  
CC actions of the polypeptide. The AvIII polynucleotide is useful as a  
CC source of probes or primers in various diagnostic assays  
XX  
SQ Sequence 88 AA;  
Query Match 98.5%; Score 462; DB 6; Length 88;  
Best Local Similarity 97.7%; Pred. No. 3.5e-45;  
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 VSGGVKVOYKNDSSAPGDNQIKPGLQVNTGSSVDLSTVTYRWFTRDGSSTLVNCD 60  
Db 1 VSGGVKVOYKNDSSAPGDNQIKPGLQVNTGSSVDLSTVTYRWFTRDGSSTLVNCD 60  
Qy 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
Db 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88

RESULT 12  
AD052317  
ID AD052317 standard; protein; 88 AA.  
XX  
AC AD052317;  
XX  
XX 15-JUL-2004 (first entry)  
XX  
XX A. cellulolyticus AvIII CBD\_III partial protein #2.  
XX  
XX Thermostable cellulase; avicelase III; AvIII; cellulose reduction;  
KW biomass degradation; ethanol formation; industrial chemical;  
KW fabric treatment; carbohydrate binding domain type III; CBD III.  
XX  
XX Acidothermus cellulolyticus.  
OS  
XX US2004038334-A1.  
XX  
XX 26-FEB-2004.  
XX  
XX 28-JUL-2001; 2001US-00917376.  
XX  
XX 28-JUL-2001; 2001US-00917376.  
XX  
XX (DING/) DING S.  
PA (ADNE/) ADNEY W S.  
PA (VINZ/) VINZANT T B.  
PA (HIMM/) HIMMEL M E.  
XX  
XX Ding S, Adney WS, Vinzant TB, Himmel ME;  
PI WPI; 2004-203224/19.  
DR  
XX Novel thermostable AvIII polypeptide of glycoside hydrolase family and  
PT isolated from Acidothermus cellulolyticus, useful for degrading treated  
PT biomass into simpler forms of carbohydrate.  
XX  
XX Claim 28; SEQ ID NO 5; 19pp; English.  
XX  
XX The invention relates to a thermostable cellulase enzyme, avicelase III  
CC (AvIII) and its nucleic acid sequence. AvIII is useful for reducing  
CC cellulose in a starting material. A thermostable AvIII peptide is useful  
CC for degrading treated biomass into simpler forms of carbohydrate, which  
CC is used in the formation of ethanol or other industrial chemicals. It is  
CC also useful for treating fabrics to remove cellulose-containing stains.  
CC The present sequence is Acidothermus cellulolyticus AvIII carbohydrate  
CC binding domain type III (CBD\_III) partial protein.  
XX  
SQ Sequence 88 AA;  
Query Match 98.5%; Score 462; DB 8; Length 88;  
Best Local Similarity 97.7%; Pred. No. 3.5e-45;  
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 VSGGVKVOYKNDSSAPGDNQIKPGLQVNTGSSVDLSTVTYRWFTRDGSSTLVNCD 60  
Db 1 VSGGVKVOYKNDSSAPGDNQIKPGLQVNTGSSVDLSTVTYRWFTRDGSSTLVNCD 60  
Qy 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
Db 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88

RESULT 13  
ABP73017  
ID ABP73017 standard; peptide; 89 AA.  
XX  
AC ABP73017;  
XX  
XX 03-JUN-2003 (first entry)  
XX  
XX Amino acid sequence of AvIII carbohydrate binding domain type III.  
DE  
XX



KW Avicelase; Aviii; glycoside hydrolase; enzyme; cellulase; biofuel;  
 KW detergent; pulp processing; paper processing; feed processing; textile;  
 KW cellulose.

XX Acidothermus cellulolyticus.

XX Key Location/Qualifiers

XX Misc-difference 89 /note= "unspecified residue"

XX WO2003012090-A2.

XX 13-FEB-2003.

XX 28-JUL-2001; 2001WO-US023818.

XX 28-JUL-2001; 2001WO-US023818.

XX (MIDE ) MIDWEST RES INST.

XX Ding S, Adney WS, Vinzant TB, Himmel ME;

XX WPI; 2003-248177/24.

XX New thermostable Aviii peptide from Acidothermus cellulolyticus, useful  
 PT for degradation of cellulose or in generating anti-Aviii antibodies for  
 PT purifying recombinant Aviii polypeptides from genetically engineered  
 PT host cells.

XX Claim 6; Page 8; 44pp; English.

XX The present sequence is derived from a thermostable avicelase, designated  
 CC Aviii. Aviii is a member of the glycoside hydrolase family of enzymes,  
 CC and is a cellulase. Aviii is useful in the conversion of biomass to  
 CC biofuels and biofuel additives. It may be useful in the production of  
 CC detergents, pulp and paper processing, food and feed processing and in  
 CC textile processes. The thermostable Aviii peptide is useful in the  
 CC degradation of cellulose, and in generating specific anti-Aviii  
 CC antibodies that are useful in purifying recombinant Aviii polypeptides  
 CC from genetically engineered host cells, in detecting Aviii polypeptide  
 CC expression, as well as a reagent tool for characterizing the molecular  
 CC actions of the polypeptide. The Aviii polynucleotide is useful as a  
 CC source of probes or primers in various diagnostic assays

XX Sequence 89 AA;

Query Match 98.5%; Score 462; DB 6; Length 89;  
 Best Local Similarity 97.7%; Pred. No. 3.5e-45;  
 Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYRYWFTRDGSSSTLVNCD 60

Db 1 VSGGVKQVYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYRYWFTRDGSSSTLVNCD 60

QY 61 WAAMCGGNIRASFGSVNPTPTADTYLQ 88

Db 61 WAAIGCGNIRASFGSVNPTPTADTYLQ 88

RESULT 14

AD052316

ID AD052316 standard; protein; 89 AA.

XX AD052316;

XX 15-JUL-2004 (first entry)

XX A. cellulolyticus Aviii CBD\_III partial protein #1.

XX Thermostable cellulase; avicelase III; Aviii; cellulose reduction;  
 KW biomass degradation; ethanol formation; industrial chemical;  
 KW fabric treatment; carbohydrate binding domain type III; CBD III.

XX

OS Acidothermus cellulolyticus.

XX Key Location/Qualifiers

XX Misc-difference 89 /note= "A string of unknown amino acids (50 or more)"

XX US2004038334-A1.

XX 26-FEB-2004.

XX 28-JUL-2001; 2001US-00917376.

XX 28-JUL-2001; 2001US-00917376.

XX (DING/) DING S.

XX (ADNEY/) ADNEY W S.

XX (VINZ/) VINZANT T B.

XX (HIMM/) HIMMEL M E.

XX Ding S, Adney WS, Vinzant TB, Himmel ME;

XX WPI; 2004-203224/19.

XX Novel thermostable Aviii polypeptide of glycoside hydrolase family and  
 PT isolated from Acidothermus cellulolyticus, useful for degrading treated  
 PT biomass into simpler forms of carbohydrate.

XX Claim 28; SEQ ID NO 4; 19pp; English.

XX The invention relates to a thermostable cellulase enzyme, avicelase III  
 CC (Aviii) and its nucleic acid sequence. Aviii is useful for reducing  
 CC cellulose in a starting material. A thermostable Aviii peptide is useful  
 CC for degrading treated biomass into simpler forms of carbohydrate, which  
 CC is used in the formation of ethanol or other industrial chemicals. It is  
 CC also useful for treating fabrics to remove cellulose-containing stains.  
 CC The present sequence is Acidothermus cellulolyticus Aviii carbohydrate  
 CC binding domain type III (CBD\_III) partial protein.

XX Sequence 89 AA;

Query Match 98.5%; Score 462; DB 8; Length 89;  
 Best Local Similarity 97.7%; Pred. No. 3.5e-45;  
 Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYRYWFTRDGSSSTLVNCD 60

Db 1 VSGGVKQVYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYRYWFTRDGSSSTLVNCD 60

QY 61 WAAMCGGNIRASFGSVNPTPTADTYLQ 88

Db 61 WAAIGCGNIRASFGSVNPTPTADTYLQ 88

RESULT 15

ABP73015

ID ABP73015 standard; protein; 957 AA.

XX AC ABP73015;

XX 03-JUN-2003 (first entry)

XX Amino acid sequence of the avicelase Aviii.

XX Avicelase; Aviii; glycoside hydrolase; enzyme; cellulase; biofuel;  
 KW detergent; pulp processing; paper processing; feed processing; textile;  
 KW cellulose.

XX Acidothermus cellulolyticus.

XX Key Location/Qualifiers

XX Misc-difference 957

XX /note= "unspecified residue encoded by N"

XX



PN WO2003012090-A2.  
XX  
PD 13-FEB-2003.  
XX  
PF 28-JUL-2001; 2001WO-US023818.  
XX  
PR 28-JUL-2001; 2001WO-US023818.  
XX  
PA (MIDE ) MIDWEST RES INST.  
XX  
PI Ding S, Adney WS, Vinzant TB, Himmel ME;  
XX  
DR WPI; 2003-248177/24.  
XX  
DR N-PSDB; ABZ77632.  
XX  
PT New thermostable Aviii peptide from Acidothermus cellulolyticus, useful  
PT for degradation of cellulose or in generating anti-Aviii antibodies for  
PT purifying recombinant Aviii polypeptides from genetically engineered  
PT host cells.  
XX  
PS Claim 2; Page 20; 44pp; English.  
XX  
CC The present sequence represents a thermostable avicelase polypeptide,  
CC designated Aviii. Aviii is a member of the glycoside hydrolase family  
CC of enzymes, and is a cellulase. Aviii is useful in the conversion of  
CC biomass to biofuels and biofuel additives. It may be useful in the  
CC production of detergents, pulp and paper processing, food and feed  
CC processing and in textile processes. The thermostable Aviii peptide is  
CC useful in the degradation of cellulose, and in generating specific anti-  
CC Aviii antibodies that are useful in purifying recombinant Aviii  
CC polypeptides from genetically engineered host cells, in detecting Aviii  
CC polypeptide expression, as well as a reagent tool for characterizing the  
CC molecular actions of the polypeptide. The Aviii polynucleotide is useful  
CC as a source of probes or primers in various diagnostic assays  
XX  
SQ Sequence 957 AA;  
Query Match 98.5%; Score 462; DB 6; Length 957;  
Best Local Similarity 97.7%; Pred. No. 6.6e-44;  
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VSGGVKQYKNDGAPGDNQIKPGLQVNTGSSVDLSTVTYRWFTRDGGSTLVYVNC 60  
Db 869 VSGGVKQYKNDGAPGDNQIKPGLQVNTGSSVDLSTVTYRWFTRDGGSTLVYVNC 928  
QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
Db 929 WAAIGCGNIRASFGSVNPATPTADTYLQ 956

Search completed: October 5, 2005, 08:03:11  
Job time : 16.8581 secs







Best Local Similarity 43.5%; Pred. NO. 1.4e-11;  
Matches 37; Conservative 14; Mismatches 33; Indels 1; Gaps 1;

QY 4 GVKVQYKNDSPAGDNQIKPGQLVNTGSSVDLSTVTYVWFTRDGGSSILVYNCWAA 63  
Db 1113 GLKVLKYNNTSASTSRPFKIVNGSSVDLSRVKIRYWTVDGDKPOSAB-CDWAQ 1171

QY 64 MGCNGIRASFGSVNPATPTADTYLQ 88  
Db 1172 IGASNVTFNFVKLTSGVSGADYYLE 1196

RESULT 3

G69593

cellulase (EC 3.2.1.4) bglC precursor - Bacillus subtilis

N:Alternate names: endo-1,4-beta-glucanase

C:Species: Bacillus subtilis

C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 09-Jul-2004

C:Accession: G69593; A26114; I40353; S24239; S49103; I39803

R:Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte

C: Bron, S.; Brouillet, S.; Bruchi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch

A.; Ehrlich, S.D.; Emerson, P.T.; Enician, K.D.; Errington, J.; Fabret, C.; Ferrati, E.

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle

lech, J.; Harwood, C.R.; Hent, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.

Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue

Y, M.; Ogawa, K.; Ogawara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,

A:Authors: Schleith, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowaka, A.; Seron

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpestra, P.; Tognoni, A.; Tosato, V.; Uchiyama,

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yanane, K.; Yasumoto, K.; Yata, K.; Yoshida, K

A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

A>Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A:Reference number: A69580; MUID:98044033; PMID:9384377

A:Accession: G69593

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-508 <KUN>

A:Cross-references: UNIPROT:P10475; UNIPROT:Q03882; GB:Z99113; GB:AL009126; NID:G2634090

R:MacKay, R.M.; Lo, A.; Willick, G.; Zuker, M.; Baird, S.; Dove, M.; Moranelli, F.; Sali

Nucleic Acids Res. 14, 9159-9170, 1986

A>Title: Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene.

A:Reference number: A26114; MUID:87066783; PMID:3024130

A:Accession: A26114

A:Molecule type: DNA

A:Residues: 10-508 <MAC>

A:Cross-references: GB:X04689; NID:G39823; PIDN:CAA28392.1; PID:G39824

A:Experimental source: strain PAP115

A>Note: part of this sequence, including the amino end of the mature form, was confirmed

R:Lindahl, V.; Aa, K.; Tronemo, A.

Antonie Van Leeuwenhoek 66, 327-332, 1994

A>Title: Nucleotide sequence of an endo-beta-1,4-glucanase gene from Bacillus subtilis

A:Reference number: I40353; MUID:95225656; PMID:7710280

A:Accession: I40353

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 10-291, 'N', 293-508 <LIN1>

A:Cross-references: EMBL:X67044; NID:G39776; PIDN:CAA47429.1; PID:G39777

R:Lindahl, V.; Aa, K.

submitted to the EMBL Data Library, June 1992

A:Reference number: S24239

A:Accession: S24239

A:Molecule type: DNA

A:Residues: 10-291, 'N', 293-508 <LIN2>

A:Cross-references: EMBL:X67044; NID:G39776; PIDN:CAA47429.1; PID:G39777

A:Experimental source: strain CK-2

R:Wolff, M.; Geczi, A.; Borris, R.

submitted to the EMBL Data Library, December 1993

A>Description: Genes encoding beta glucan-hydrolyzing enzymes in Bacillus subtilis: cons

A:Reference number: S49103

A:Accession: S49103

A:Molecule type: DNA

A:Residues: 10-508 <WOL>

A:Cross-references: EMBL:Z29076; NID:G509266; PIDN:CAA2317.1; PID:G509267

R:Seo, Y.S.; Lee, Y.H.; Pek, U.H.; Kang, H.

Korean J. Microbiol. 24, 236-242, 1986

A>Title: Analysis on the nucleotide sequence of the signal region of Bacillus subtilis

A:Reference number: I39803

A:Accession: I39803

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 10-13, 'V', 15, 'N', 17-18, 'V', 20-21, 'P', 23, 'A', 25-26, 'AI', 29-31, 'P', 33, 'PO', 36,

A:Cross-references: GB:M38634; NID:G142657; PIDN:AAA22300.1; PID:G142658

A:Experimental source: strain ATCC 6633

C:Comment: The low molecular weight of the mature protein suggests carboxyl-terminal pro

C:Genetics:

A:Gene: bglC

C:Function:

A>Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel

A:Pathway: cellulose degradation

C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation

F:1-38/Domain: signal sequence #status predicted <SIG>

F:39-508/Product: cellulase #status predicted <MAT>

Query Match 40.5%; Score 190; DB 2; Length 508;  
Best Local Similarity 40.7%; Pred. No. 4.2e-12;  
Matches 35; Conservative 21; Mismatches 26; Indels 4; Gaps 2;

QY 4 GVKVQYKNDSPAGDNQIKPGQLVNTGSSVDLSTVTYVWF-TRDGGSSILVYNCWAA 62  
Db 362 GISVQYKAGDGRVNSNQIRPQLHIKNNGNATVDLKDVTARYWYKAKNKGQ---FDCDYA 418

QY 63 AMGCNGIRASFGSVNPATPTADTYLQ 88  
Db 419 QIGCGNVTHKFTLHKPKQGADTYLE 444

RESULT 4

A26874

cellulase (EC 3.2.1.4) precursor - Bacillus subtilis (strain DLG)

N:Alternate names: endo-1,4-beta-glucanase

C:Species: Bacillus subtilis

C>Date: 08-Mar-1989 #sequence\_revision 08-Mar-1989 #text\_change 09-Jul-2004

C:Accession: A26874; B26874

R:Robson, L.M.; Chambliss, G.H.

J. Bacteriol. 169, 2017-2025, 1987

A>Title: Endo-beta-1-4-glucanase gene of Bacillus subtilis DLG.

A:Reference number: A26874; MUID:87194581; PMID:3106328

A:Accession: A26874

A:Molecule type: DNA

A:Residues: 1-508 <ROB1>

A:Cross-references: UNIPROT:P07983; GB:M16185; NID:G143007; PIDN:AAA22496.1; PID:G143007

A:Experimental source: strain DLG

A:Accession: B26874

A:Molecule type: protein

A:Residues: 39-53 <ROB2>

A:Experimental source: strain DLG

A>Note: the authors believe Met-1 and Met-2 may be alternate initiators

C:Comment: The low molecular weight of the mature protein suggests carboxyl-terminal pro

C:Function:

A>Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel

A:Pathway: cellulose degradation

C:Keywords: alternative initiators; extracellular protein; glycosidase; hydrolase; poly

F:1-38/Domain: (or 2-38) signal sequence #status predicted <SIG>

Query Match 40.5%; Score 190; DB 2; Length 508;  
Best Local Similarity 43.0%; Pred. No. 4.2e-12;  
Matches 37; Conservative 17; Mismatches 28; Indels 4; Gaps 2;

QY 4 GVKVQYKNDSPAGDNQIKPGQLVNTGSSVDLSTVTYVWF-TRDGGSSILVYNCWAA 62  
Db 362 GISVQYKAGDGRVNSNQIRPQLHIKNNGNATVDLKDVTARYWYKAKNKGQ---FDCDYA 418

QY 63 AMGCNGIRASFGSVNPATPTADTYLQ 88



Db 419 QMCGNLTHKFTVTLHKPKQGADTYLE 444  
 RESULT 5  
 JN0111  
 Cellulase (EC 3.2.1.4) precursor - Bacillus subtilis (strain BSE616)  
 N:Alternate names: endo-1,4-beta-glucanase  
 C:Species: Bacillus subtilis  
 C:Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 09-Jul-2004  
 C:Accession: JN0111  
 R:Park, S.H.; Kim, H.K.; Pack, M.Y.  
 Agric. Biol. Chem. 55, 441-448, 1991  
 A:Title: Characterization and structure of the cellulase gene of Bacillus subtilis BSE616  
 A:Reference number: JN0111; MUID:91299280; PMID:1368694  
 A:Accession: JN0111  
 A:Molecule type: DNA  
 A:Residues: 1-499 <PAR>  
 A:Cross-references: UNIPROT:P23549; GB:D01057; NID:9216387; PIDN:BAA00859.1; PID:d100132  
 A:Note: The authors translated the codon ATA for residue 102 as Tyr  
 C:Comment: The low molecular weight of the mature protein suggests carboxyl-terminal processing.  
 C:Function:  
 A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cellulose  
 A:Pathway: cellulose degradation  
 C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation  
 F:1-29/Domain: signal sequence #status predicted <SIG>  
 Query Match 40.1%; Score 188; DB 2; Length 499;  
 Best Local Similarity 40.0%; Pred. No. 6.6e-12;  
 Matches 34; Conservative 18; Mismatches 31; Indels 2; Gaps 1;  
 Qy 4 GVKVQYKNDSPAGDNOIKPGLQLVNTGSSSVLDSTVTYRVYFTRDGGSTLVYNCDDAA 63  
 Db 353 GISVQYRAGDSWMSNQIRPOLQKNGNTTVDLKDVTARYWYNAKNGQNV--DCDYAQ 410  
 Qy 64 MCGGNIRASFGSVNPAFTADTYLQ 88  
 Db 411 LGCGNVTKFVTLHKPKQGADTYLE 435  
 RESULT 6  
 A48954  
 mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Caldocellum saccharolyticum  
 N:Alternate names: beta-mannanase  
 C:Species: Caldocellum saccharolyticum  
 C:Date: 19-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004  
 C:Accession: A48954; B43745  
 R:Gibbs, M.D.; Saul, D.J.; Luthi, E.; Bergquist, P.L.  
 Appl. Environ. Microbiol. 58, 3864-3867, 1992  
 A:Title: The beta-mannanase from "Caldocellum saccharolyticum" is part of a multidomain  
 A:Reference number: A48954; MUID:93119139; PMID:1476429  
 A:Accession: A48954  
 A:Status: preliminary  
 A:Molecule type: nucleic acid  
 A:Residues: 1-1331 <GIB>  
 A:Cross-references: UNIPROT:P22533; GB:L01257; NID:9144290; PIDN:AAA71887.1; PID:g144291  
 A:Note: sequence extracted from NCBI backbone (NCBI:121576, NCBI:121577)  
 R:Luethi, E.; Bhana Jasmal, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.  
 Appl. Environ. Microbiol. 57, 694-700, 1991  
 A:Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene coding  
 A:Reference number: A43745; MUID:91247819; PMID:2039230  
 A:Accession: B43745  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-337, "PPRQHOHRQ" <LUE>  
 A:Cross-references: EMBL:M36063; NID:g144292; PIDN:AAA72861.1; PID:g144294  
 A:Note: the authors translated the codon CAC for residue 262 as Glu  
 A:Note: this sequence has been revised in reference A48954  
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation  
 Query Match 39.8%; Score 186.5; DB 2; Length 1331;  
 Best Local Similarity 42.5%; Pred. No. 2.7e-11;  
 Matches 37; Conservative 18; Mismatches 31; Indels 1; Gaps 1;



```

Query Match      38.9%; Score 182.5; DB 2; Length 1039;
Best Local Similarity 41.9%; Pred. No. 5.4e-11;
Matches 36; Conservative 18; Mismatches 31; Indels 1; Gaps 1
Qy      3  GGKVKVQKNNDSPCDNQIKPGLQVNTGSSSVDLSTVTYRWFYRDGSSLTIVYNCDA 62
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      420  GQIKVLANKETNSTTIRPWLKVNSSGSSIDLSTVIRYWTVDGERAQSAVS-DWA 478

Qy      63  AMCGCNTRASFGSVNPATPTADTYLQ 88
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      479  QIGASNVTFKFKVLSSSVSGADYYLE 504

```

```

RESULT 11
A41897
cellulase homolog - Bacillus lautus (fragment)
C/Species: Bacillus lautus
C/Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 09-Jul-2004
C/Accession: A41897; S27498
R/Hansen, C.K.; Diderichsen, B.; Jorgensen, P.L.
J. Bacteriol. 174, 3522-3531, 1992
A/Title: celiA from Bacillus lautus PL236 encodes a novel cellulose-binding endonuclease
A/Reference number: A41897; MUID:92276330; PMID:1592807
A/Accession: A41897
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-145 <HAN>
A/Cross-references: UNIPROT:P29718; EMBL:M76588; NID:G142661; PIDN:AAA22302.1.1
A/Experimental source: PL236
A/Note: sequence extracted from NCBI backbone (NCBIP:104604)

Query Match 38.5%; Score 180.5; DB 2; Length 145;
Best Local Similarity 43.9%; Pred. No. 1.1e-11;
Matches 36; Conservative 18; Mismatches 25; Indels 3; Gaps 2

Qy 7 VQYKNDSAPGDNIKPGQLQVNTGSSVDLSTVTVRYWFTRGDGSSTLVYNCDMAAGC 66
Db 1 LQYRADTNAADNQIKFSFNKNGTSAVDLSTLKIRYFTKGSAAVNGW-IDWAQLGG 59

Qy 67 GNIRASFGSNPATPTADTYLQ 88
Db 60 SNIQISFG--NHTGTNSDTYVE 79

RESULT 12
I40548
bifunctional cellulase precursor - Bacillus sp.
C/Species: Bacillus sp.
C/Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 09-Jul-2004
C/Accession: I40548
R/Han, S.J.; Yoo, Y.J.; Kang, H.S.
J. Biol. Chem. 270, 26012-26019, 1995
A/Title: Characterization of a bifunctional cellulase and its structural gene
A/Reference number: I40548; MUID:96029707; PMID:7592793
A/Accession: I40548
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-486 <RES>
A/Cross-references: UNIPROT:Q45430; EMBL:U27084; NID:9857575; PIDN:AAC43478.1.1
C/Genetics:
A/Gene: cel

Query Match 38.4%; Score 180; DB 2; Length 486;
Best Local Similarity 38.8%; Pred. No. 4.4e-11;
Matches 33; Conservative 18; Mismatches 32; Indels 2; Gaps 1

Qy 4 GVKYQYKNDSAPGDNIKPGQLQVNTGSSVDLSTVTVRYWFTRGDGSSTLVYNCDMAA 63
Db 353 GISVQYRAGDGSMMNSQIRPQLQKNGNTTVDLKDVTARYWYNAKNGQL--DCDYE 410

Qy 64 MGCNGIRASFGSNPATPTADTYLQ 88

```



Db 411 LGCNVSHVTTLHKPKQADTYLE 435

RESULT 13

B41897  
cellulase (EC 3.2.1.4) - Bacillus lautus  
N:Alternate names: endo-1,4-beta-glucanase  
C:Species: Bacillus lautus  
C:Date: 10-Mar-1994 #sequence\_revision 10-Mar-1994 #text\_change 09-Jul-2004  
C:Accession: B41897; S27499  
R:Hansen, C.K.; Diderichsen, B.; Jorgensen, P.L.  
J. Bacteriol. 174, 3522-3531, 1992  
A:Title: cclA from Bacillus lautus PL236 encodes a novel cellulose-binding endo-beta-1,4  
A:Reference number: A41897; MUID:92276330; PMID:1592807  
A:Accession: B41897  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-700 <HAN>  
A:Cross-references: UNIPROT:P29719; EMBL:M76588; NID:G142661; PIDN:AAA22303.1; PID:G14266  
A:Experimental source: PL236  
A:Note: sequence extracted from NCBI backbone (NCBIP:104605)  
C:Function:  
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce  
A:Pathway: cellulose degradation  
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 36.7%; Score 172; DB 2; Length 700;  
Best Local Similarity 44.9%; Pred. No. 4.4e-10;  
Matches 40; Conservative 18; Mismatches 27; Indels 4; Gaps 3;

QY 1 VSGGVKVOYKND-SAPGDNQIKPGLQVNTGSSVDLSTVTYRWFTRDGGSSTLVYNC 59  
Db 549 VNSDLVQYKQDRNATDNQIKFNIQKGTSPVLSLTLRYFTKD-SSAANGWI 607

QY 60 DNAAMCGNIRASFGSVNPTPTADTYLQ 88  
Db 608 DWAKUGGSIQISFGNHGA--DSDIYAE 634

RESULT 14

A47704  
endoglucanase I (EC 3.2.1.-) Celli - Clostridium thermocellum  
C:Species: Clostridium thermocellum  
C:Date: 19-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004  
C:Accession: A47704  
R:Hazlewood, G.P.; Davidson, K.; Laurie, J.I.; Huskisson, N.S.; Gilbert, H.J.  
J. Gen. Microbiol. 139, 307-316, 1993  
A:Title: Gene sequence and properties of Celli, a family E endoglucanase from Clostridium  
A:Reference number: A47704; MUID:93171873; PMID:8436949  
A:Accession: A47704  
A:Status: preliminary  
A:Molecule type: DNA; protein  
A:Residues: 1-879 <HAZ>  
A:Cross-references: UNIPROT:Q02934; GB:I04735; NID:G144807; PIDN:AAA20892.1; PID:G144808  
A:Note: sequence extracted from NCBI backbone (NCBIN:125637, NCBIP:125638)  
C:Keywords: glycosidase; hydrolase

Query Match 35.3%; Score 165.5; DB 2; Length 879;  
Best Local Similarity 33.0%; Pred. No. 2.7e-09;  
Matches 29; Conservative 26; Mismatches 32; Indels 1; Gaps 1;

QY 1 VSGGVKVOYKND-SAPGDNQIKPGLQVNTGSSVDLSTVTYRWFTRDGGSSTLVYNC 60  
Db 736 IKGEVLQYANGAGTNSINPRFKIINNGTKAINLSVDKIRYYTYKGGASQNF-CD 794

QY 61 WAAMCGNIRASFGSVNPTPTADTYLQ 88  
Db 795 WSSAGNSNVGTGNFNLSPKEGADTCLE 822

RESULT 15

S36859  
cipA protein - Clostridium thermocellum

N:Alternate names: probable cellulosome protein large chain SL  
C:Species: Clostridium thermocellum  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 09-Jul-2004  
C:Accession: S36859; S33527; S25767; S28659; T18261  
R:Gerrigross, U.T.; Demain, A.L.  
submitted to the EMBL Data Library, January 1993  
A:Reference number: S36859  
A:Accession: S36859  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1854 <GER>  
A:Cross-references: UNIPROT:Q06851; EMBL:L08665  
R:Gerrigross, U.T.; Romaniec, M.P.M.; Kobayashi, T.; Huskisson, N.S.; Demain, A.L.  
Mol. Microbiol. 8, 325-334, 1993  
A:Title: Sequencing of a Clostridium thermocellum gene (cipA) encoding the cellulosomal  
A:Reference number: S33527; MUID:93302508; PMID:8316083  
A:Accession: S33527  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1615, 1617-1854 <GE2>  
A:Cross-references: EMBL:L08665  
R:Fujino, T.; Beguin, P.; Aubert, J.P.  
FEMS Microbiol. Lett. 94, 165-170, 1992  
A:Title: Cloning of a Clostridium thermocellum DNA fragment encoding polypeptides that  
A:Reference number: S25767  
A:Accession: S25767  
A:Molecule type: DNA  
A:Residues: 1216-1232, 'Y', 1234-1241, 'T', 1243-1319, 'R', 1321-1615, 1617-1854 <FUJ>  
A:Cross-references: EMBL:X67406  
R:Bequin, P.  
submitted to the EMBL Data Library, August 1992  
A:Reference number: S28659  
A:Accession: S28659  
A:Molecule type: DNA  
A:Residues: 1216-1232, 'Y', 1234-1241, 'T', 1243-1319, 'R', 1321-1615, 1617-1768, 'R', 1770-1854  
A:Cross-references: EMBL:X67406  
R:Fujino, T.; Beguin, P.; Aubert, J.P.  
J. Bacteriol. 175, 1891-1899, 1993  
A:Title: Organization of a Clostridium thermocellum gene cluster encoding the celluloso  
e.  
A:Reference number: Z18847; MUID:93209931; PMID:8458832  
A:Accession: T18261  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1821-1854 <FU2>  
A:Cross-references: EMBL:X67506; NID:G296879; PID:G296880; PIDN:CAA47840.1  
C:Genetics:  
A:Gene: cipA

Query Match 35.2%; Score 165; DB 2; Length 1854;  
Best Local Similarity 36.1%; Pred. No. 6.7e-09;  
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKVOYKND-SAPGDNQIKPGLQVNTGSSVDLSTVTYRWFTRDGGSSTLVYNC 60  
Db 365 VSGNLKVEFYNPSPTTNSINPQFKVTNTGSSAIDLSKLTLYRYTYVDGQKDTFW-CD 423

QY 61 WAAM-----GCGNIRASFGSVNPTPTADTYLQ 88  
Db 424 HAAIIGNSYNGITNSVKGTFFVKMSSTNNADTYLE 460

Search completed: October 5, 2005, 07:57:10  
Job time : 3.44184 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 07:44:57 ; Search time 10.9413 Seconds  
(without alignments)  
4118.611 Million cell updates/sec

Title: US-09-917-376-5  
Perfect score: 469  
Sequence: 1 VSGGVKQYKKNDSAPGDNQ.....IRASFGSVNPATPTADTYLQ 88

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt\_03:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	257	54.8	741	Q82QF2	Q82qf2 streptomyce
2	228.5	48.7	616	Q7X2N2	Q7x2n2 thermomonos
3	225.5	48.1	170	Q9RFK6	Q9rfk6 caldibacill
4	216.5	46.2	930	Q9RFK5	Q9rfk5 caldibacill
5	214.5	45.7	921	Q918L8	Q918l8 caldibacill
6	209.5	44.7	1091	Q8KFF7	Q8kff7 paenibacill
7	204.5	43.6	997	Q9Z4I1	Q9z4i1 bacillus sp
8	194.5	41.5	1751	Q9AQG4	Q9aqg4 caldicellul
9	192.5	41.0	1000	Q24820	Q24820 thermophil
10	192.5	41.0	1770	Q9X3P5	Q9x3p5 caldicellul
11	191.5	40.8	261	Q9AQG7	Q9aqg7 caldicellul
12	191.5	40.8	1426	Q9X3P6	Q9x3p6 caldicellul
13	191	40.7	1711	P96311	P96311 anaerocellu
14	190.5	40.6	996	Q9AQH0	Q9aqh0 caldicellul
15	190.5	40.6	1779	Q52374	Q52374 caldicellul
16	190	40.5	499	GUN1_BACSU	P107983 bacillus su
17	190	40.5	499	GUN2_BACSU	P107983 bacillus su
18	190	40.5	499	Q93TJ6	Q93tj6 bacillus su
19	190	40.5	508	Q93LD0	Q93ld0 bacillus su
20	190	40.5	508	Q70K05	Q70k05 bacillus am
21	188	40.1	499	GUN3_BACSU	P23349 bacillus su
22	187	39.9	499	Q52731	O52731 bacillus sp
23	187	39.9	499	Q6Y4Q1	Q6y4q1 bacillus su
24	186.5	39.8	1331	MANB_CALSA	P22533 caldocellum
25	186.5	39.8	1742	GUNA_CALSA	P22534 caldocellum
26	186	39.7	499	Q45532	Q45532 bacillus su
27	186	39.7	501	Q83012	Q83012 bacillus sp
28	182.5	38.9	1039	GUNB_CALSA	P10474 c endogluc
29	180.5	38.5	145	YCEA_PAEIA	P29718 paenibacill
30	180	38.4	486	Q45430	Q45430 bacillus sp
31	177	37.7	499	Q8RPQ6	Q8rpq6 bacillus am

RESULT 1

ID	Q82QF2	PRELIMINARY;	PRT;	741 AA.
AC	Q82QF2;			
DT	01-JUN-2003 (TrEMBLrel. 24, Created)			
DT	01-JUN-2003 (TrEMBLrel. 24, Last sequence update)			
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)			
DE	Putative cellulose 1,4-beta-cellobiosidase.			
GN	Names=guxA1; OrderedLocustNames=SAV557;			
OS	Streptomyces avermitilis.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Streptomycineae; Streptomycetaceae; Streptomyces.			
OX	NCBI_TaxID=33903;			
RN	[1]			
RC	SEQUENCE FROM N.A.			
RP	STRAIN=MA-4680;			
RX	MEDLINE=21477403; PubMed=11572948; DOI=10.1073/pnas.211433198;			
RA	Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,			
RA	Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonce T.,			
RA	Kikuchi H., Shiba T., Sakaki Y., Hattori M.;			
RT	"Genome sequence of an industrial microorganism Streptomyces			
RT	avermitilis; deducing the ability of producing secondary			
RT	metabolites."			
RL	Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).			
RN	[2]			
RC	SEQUENCE FROM N.A.			
RP	STRAIN=MA-4680;			
RX	MEDLINE=22608306; PubMed=12692562;			
RA	Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,			
RA	Sakaki Y., Hattori M., Omura S.;			
RT	"Complete genome sequence and comparative analysis of the industrial			
RT	microorganism Streptomyces avermitilis."			
RL	Nat. Biotechnol. 21:526-531(2003).			
CC	-I- SIMILARITY: Contains 1 fibronectin type III domain.			
EMBL	AP005023; BAC68267.1; ..			
HSSP	P20533; 1K85			
GO	GO:0004553; P:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.			
GO	GO:0005975; P:carbohydrate metabolism; IEA.			
InterPro	IPR001956; CBD_3.			
InterPro	IPR008965; Cellul bind.			
InterPro	IPR003961; FN_III-like.			
InterPro	IPR008957; FN_III-like.			
InterPro	IPR01524; Glyco_hydro_6.			
InterPro	IPR011253; Glyco_hydro_6-lk.			
Pfam	PF00942; CBM_3; 1.			
Pfam	PF00041; fn3; 1.			
Pfam	PF01341; Glyco_hydro_6; 1.			
PRINTS	PR00733; GLHYDRLASE6.			
ProDom	PD001947; CBD_3; 1.			
ProDom	PD003733; Glyco_hydro_6; 1.			
SMART	SM00060; FN3; 1.			
PROSITE	PS00853; FN3; 1.			
PROSITE	PS00655; GLYCOSYL_HYDROL_F6_1; 1.			
KW	Complete proteome.			



```
SQ SEQUENCE 741 AA; 77396 MW; 79404B40B2B4A7AF CRC64;
Query Match 54.8%; Score 257; DB 2; Length 741;
Best Local Similarity 56.3%; Pred. No. 1.5e-19;
Matches 49; Conservative 13; Mismatches 25; Indels 0; Gaps 0;

Qy 2 SGGVQVQYKKNDSAPGDNQIKPGLQLVNTGSSVLDLSTVTVRYWFTRDGSSSTLVVNCWD 61
Db 591 SGGKLVLYKKNDSATDNRPLGLRIVNTGSSILDSKVTARYFTRDGSSPTVNAWCY 650

Qy 62 AAMGCNIRASFGSVNPATPTADTYLQ 88
Db 651 AAVGCSNVLKVVPLITTPVFGADAYLE 677

RESULT 2
Q7X2N2 PRELIMINARY; PRT; 616 AA.
AC Q7X2N2;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Endoglucanase.
GN Name=cel5B;
OS Thermomonospora fusca.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptosporangineae; Nocardiopsaceae; Thermobifida.
OX NCBI_TaxID=2021;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TW51;
RA Posta K., Beki E., Kukolya J., Hornok L.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY298814; AAP56348.1; -.
DR HSP; Q45996; I043.
DR GO; GO:0004553; F:carbohydrate metabolism; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00942; CBM 3; 1.
DR Pfam; PF00150; Cellulase; 1.
DR ProDom; PD001947; CBD 3; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN 1.
SQ SEQUENCE 616 AA; 67701 MW; 24FFC1E1A3F5639 CRC64;

Query Match 48.7%; Score 228.5; DB 2; Length 616;
Best Local Similarity 52.3%; Pred. No. 1.7e-16;
Matches 45; Conservative 16; Mismatches 22; Indels 3; Gaps 2;

Qy 2 SGGVQVQYKKNDSAPGDNQIKPGLQLVNTGSSVLDLSTVTVRYWFTRDGSSSTLVVNCWD 61
Db 471 TGALEVYRYNLSAADDSQIAPGLRLVNTGSSVLDLADVIHYFTNEPG-GTLQFTCDW 529

Qy 62 AAMGCNIRASFGSVNPATPTADTYL 87
Db 530 AQVGCANVNASTSL--SAPGADTSL 553

RESULT 3
Q9RFK6 PRELIMINARY; PRT; 170 AA.
AC Q9RFK6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Caldicellulosum cellulovorans.
OC Bacteria; Firmicutes; Bacillales; Alicyclobacillaceae; Caldicellulosum.
OX NCBI_TaxID=74586;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20120520; PubMed=10653733; DOI=10.1128/AEM.66.2.664-670.2000;
RX
```

```
RA Sunna A., Gibbs M.D., Chin C.W.J., Nelson P.J., Bergquist P.L.;
RT "A gene encoding a novel multidomain beta-1,4-mannanase from
RT Caldicellulosum cellulovorans and action of the recombinant enzyme on
RT kraft pulp.";
RL Appl. Environ. Microbiol. 66:664-670 (2000).
DR EMBL; AF163837; AAF22273.1; -.
DR HSP; Q06851; INBC.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul_bind.
DR Pfam; PF00942; CBM 3; 1.
DR ProDom; PD001947; CBD 3; 1.
DR KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 170 AA; 18493 MW; 7AC9D33F44E3A0B4 CRC64;

Query Match 48.1%; Score 225.5; DB 2; Length 170;
Best Local Similarity 47.7%; Pred. No. 8.3e-17;
Matches 41; Conservative 17; Mismatches 27; Indels 1; Gaps 1;

Qy 3 GGVQVQYKKNDSAPGDNQIKPGLQLVNTGSSVLDLSTVTVRYWFTRDGSSSTLVVNCWD 62
Db 20 GSVVQYRAADTNAGDNQIKPGLRIVNTGSSVPLSELTIWYTYD-GDKPQVNCWD 78

Qy 63 AAMGCNIRASFGSVNPATPTADTYLQ 88
Db 79 QVGCNVRGSPVFKLSTGRTGADYIE 104

RESULT 4
Q9RFK5 PRELIMINARY; PRT; 930 AA.
AC Q9RFK5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Multidomain beta-1,4-mannanase precursor.
GN Name=mana;
OS Caldicellulosum cellulovorans.
OC Bacteria; Firmicutes; Bacillales; Alicyclobacillaceae; Caldicellulosum.
OX NCBI_TaxID=74586;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20120520; PubMed=10653733; DOI=10.1128/AEM.66.2.664-670.2000;
RA Sunna A., Gibbs M.D., Chin C.W.J., Nelson P.J., Bergquist P.L.;
RT "A gene encoding a novel multidomain beta-1,4-mannanase from
RT Caldicellulosum cellulovorans and action of the recombinant enzyme on
RT kraft pulp.";
RL Appl. Environ. Microbiol. 66:664-670 (2000).
DR EMBL; AF163837; AAF22274.1; -.
DR HSP; Q92F13; 3MAN.
DR GO; GO:0019028; C:Viral capsid; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR004302; Chitin_binding_3.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00942; CBM 3; 2.
DR Pfam; PF00150; Cellulase; 1.
DR Pfam; PF03067; Chitin_bind_3; 1.
DR ProDom; PD001947; CBD 3; 2.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN 1.
KW Signal.
FT CHAIN 1 33 Potential.
FT SIGNAL 34 930 multidomain beta-1,4-mannanase.
SQ SEQUENCE 930 AA; 101576 MW; 0086638D54D1A2CC CRC64;

Query Match 46.2%; Score 216.5; DB 2; Length 930;
Best Local Similarity 46.5%; Pred. No. 5.9e-15;
Matches 40; Conservative 17; Mismatches 28; Indels 1; Gaps 1;

Qy 3 GGVQVQYKKNDSAPGDNQIKPGLQLVNTGSSVLDLSTVTVRYWFTRDGSSSTLVVNCWD 62
```



RESULT 6	
Q8KKF7	PRELIMINARY; PRT; 1091 AA.
ID	Q8KKF7
AC	Q8KKF7;
DT	01-OCT-2002 (TrEMBLrel. 22, Created)
DT	01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE	Cellulose 1,4-beta-cellobiosidase precursor (EC 3.2.1.91).
GN	Name=cel48C;
OS	<i>Paenibacillus</i> sp. BP-23.



```
DR InterPro; IPR008957; FN III-like.
DR InterPro; IPR01701; Glyco_hydro_9.
DR InterPro; IPR008928; Glyco_trans_6hp.
DR InterPro; IPR002345; Lipocalin.
DR Pfam; PF00942; CBM_3; 2.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF00759; Glyco_hydro_9; 1.
DR ProDom; PD001947; CBD_3; 2.
DR SMART; SMO0060; FN3; 2.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_2; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
KW Glycosidase; Hydrolase; Signal.
FT SIGNAL 1 34 Potential.
SQ SEQUENCE 997 AA; 106927 MW; F20CB119D6410629 CRC64;

Query Match 43.6%; Score 204.5; DB 2; Length 997;
Best Local Similarity 43.7%; Pred. No. 1.3e-13;
Matches 38; Conservative 21; Mismatches 27; Indels 1; Gaps 1;

QY 2 SGGVKQYKNNDSAPGDNQIKPGLQVNTGSSVDLSVTYVYWFTRDGGSSTLVYNCDW 61
Db 846 TGTLEQYRSGSGSSNAVTPQFNKNTGQAIDLSTVKIRYFTKD-GTELSFWCDY 904

QY 62 AAMGCGNIRASFGSVNPAATPTADTYLQ 88
Db 905 AQVGSANVQGMFVAVNPANPKGATDYVE 931

RESULT 8
Q9AQG4 PRELIMINARY; PRT; 1751 AA.
AC Q9AQG4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Cell.
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
RT thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
DR EMBL; AF078042; AAK06394.1; -.
DR HSSP; P37700; 1G87.
DR GO; GO:0004553; P:hydrolase activity, hydrolyzing O-glycosyl . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul. bind.
DR InterPro; IPR001701; Glyco_hydro_9.
DR InterPro; IPR008928; Glyco_trans_6hp.
DR Pfam; PF00942; CBM_3; 4.
DR Pfam; PF00759; Glyco_hydro_9; 1.
DR ProDom; PD001947; CBD_3; 4.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
SQ SEQUENCE 1751 AA; 192176 MW; 60178CBF3C00BE95 CRC64;

Query Match 41.5%; Score 194.5; DB 2; Length 1751;
Best Local Similarity 44.7%; Pred. No. 3.3e-12;
Matches 38; Conservative 14; Mismatches 32; Indels 1; Gaps 1;

QY 4 GVKVQYKNNDSAPGDNQIKPGLQVNTGSSVDLSVTYVYWFTRDGGSSTLVYNCDWAA 63
Db 4 GVKVQYKNNDSAPGDNQIKPGLQVNTGSSVDLSVTYVYWFTRDGGSSTLVYNCDWAA 63

Db 678 GVKVLYKNNETSASTGSIRPMFKIVNGSSVDLSRVKIRYWTVDGDKPQSAV-CDWAQ 736
QY 64 MCGNIRASFGSVNPAATPTADTYLQ 88
Db 737 IGASNVTFNFKVLSGSGVADYYLE 761

RESULT 9
Q24820 PRELIMINARY; PRT; 1000 AA.
AC Q24820;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Beta-glucanase.
OS thermophilic anaerobe NA10.
OC Bacteria.
OX NCBI_TaxID=67756;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NA10;
RA Miyake K., Machida Y., Hattori K., Iijima S.;
RL Submitted (Oct-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to cellulase family F (family 10 of glycosyl
CC hydrolases).
DR EMBL; AB008029; BAA22939.1; -.
DR HSSP; Q06851; 1NBC.
DR GO; GO:0004553; P:hydrolase activity, hydrolyzing O-glycosyl . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001589; Actbind_actnin.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul. bind.
DR InterPro; IPR001000; Glyco_hydro_10.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00942; CBM_3; 1.
DR Pfam; PF00331; Cellulase; 1.
DR PRINTS; PR00134; GLHYDLASE10.
DR ProDom; PD001947; CBD_3; 1.
DR SMART; SM00633; Glyco_10; 1.
DR PROSITE; PS00019; ACTININ_1; UNKNOWN_1.
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 1000 AA; 113264 MW; B9F659A56A752C6B CRC64;

Query Match 41.0%; Score 192.5; DB 2; Length 1000;
Best Local Similarity 43.7%; Pred. No. 2.9e-12;
Matches 38; Conservative 17; Mismatches 31; Indels 1; Gaps 1;

QY 2 SGGVKQYKNNDSAPGDNQIKPGLQVNTGSSVDLSVTYVYWFTRDGGSSTLVYNCDW 61
Db 372 SQGIKLYANKETNSTNTIREPLKVVNTGSSSIDLSRVTIKYWTYVDGDKAQSASV-DW 430
QY 62 AAMGCGNIRASFGSVNPAATPTADTYLQ 88
Db 431 AQIGASNVTFNFKVLSGSGVADYYLE 457

RESULT 10
Q9X3P5 PRELIMINARY; PRT; 1770 AA.
AC Q9X3P5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE XynA.
GN Name=xynA;
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
```



```
RP SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1;
RA MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RA "Multidomain and multifunctional glycosyl hydrolases from the extreme
RT thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
CC -!- SIMILARITY: Belongs to cellulase family F (family 10 of glycosyl
CC hydrolases).
DR EMBL; AF078737; AAD30363.1; -.
DR HSSP; Q06851; INBC.
DR GO; GO:0030246; P:carbohydrate binding; IEA.
DR GO; GO:0004553; P:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR006584; CBD_IV.
DR InterPro; IPR005084; CBM_6.
DR InterPro; IPR003305; CBM_CenC.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR008979; Gal_bind_like.
DR InterPro; IPR001000; Glyco_hydro_10.
DR InterPro; IPR006710; Glyco_hydro_43.
DR Pfam; PF00942; CBM_3; 3.
DR Pfam; PF02018; CBM_4_9; 2.
DR Pfam; PF03422; CBM_6; 1.
DR Pfam; PF00331; Glyco_hydro_10; 1.
DR Pfam; PF04616; Glyco_hydro_43; 1.
DR PRINTS; PR00134; GLHYDRLASE10.
DR ProDom; PD001947; CBD_3; 3.
DR SMART; SM00606; CBD_IV; 1.
DR SMART; SM00633; Glyco_10; 1.
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 1770 AA; 8BAF1937D4926C92 CRC64;

Query Match 41.0%; Score 192.5; DB 2; Length 1770;
Best Local Similarity 43.5%; Pred. No. 5.6e-12;
Matches 37; Conservative 16; Mismatches 31; Indels 1; Gaps 1;

QY 4 GVKVQYKNDSAPGDNQIKPGLQVLNTGSSVDLSVTYVRYWFTRDGSGSTLVYNCDWAA 63
Db 1104 GLKVLVYKNNETSASTSSIRPFKIVNGSSVDLSRVKIRYWTVDGDKPQSAV-CDWAQ 1162

QY 64 MCGCNIRASFGSVNPATPTADTYLQ 88
Db 1163 IGASNVTFNFVKLSGSGADYYLE 1187

RESULT 11
Q9AQG7 PRELIMINARY; PRT; 261 AA.
AC Q9AQG7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Glycosyl hydrolase 6 (Fragment).
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RA "Multidomain and multifunctional glycosyl hydrolases from the extreme
RT thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
DR EMBL; AF078040; AAK06391.1; -.
DR HSSP; Q06851; INBC.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
```

```
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul_bind.
DR Pfam; PF00942; CBM_3; 1.
DR ProDom; PD001947; CBD_3; 1.
KW Hydrolase.
FT NON_TER 1 1
FT NON_TER 261 261
SQ SEQUENCE 261 AA; 28759 MW; 4771744A26A6E04 CRC64;

Query Match 40.8%; Score 191.5; DB 2; Length 261;
Best Local Similarity 43.5%; Pred. No. 7.7e-13;
Matches 37; Conservative 15; Mismatches 32; Indels 1; Gaps 1;

QY 4 GVKVQYKNDSAPGDNQIKPGLQVLNTGSSVDLSVTYVRYWFTRDGSGSTLVYNCDWAA 63
Db 118 GLKVLVYKNNETSASTSSIRPFKIVNGSSVDLSRVKIRYWTVDGDKPQSAV-CDWAQ 176

QY 64 MCGCNIRASFGSVNPATPTADTYLQ 88
Db 177 IGASNVTFNFVKLSGSGADYYLE 201

RESULT 12
Q9X3P6 PRELIMINARY; PRT; 1426 AA.
AC Q9X3P6;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE CellB.
GN Name=cellB;
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1;
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RA "Multidomain and multifunctional glycosyl hydrolases from the extreme
RT thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
CC -!- SIMILARITY: Belongs to cellulase family F (family 10 of glycosyl
CC hydrolases).
DR EMBL; AF078737; AAD30364.1; -.
DR HSSP; Q06851; INBC.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001589; Actbind_actnin.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR001000; Glyco_hydro_10.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00942; CBM_3; 3.
DR Pfam; PF00150; Cellulase; 1.
DR Pfam; PF00331; Glyco_hydro_10; 1.
DR PRINTS; PR00134; GLHYDRLASE10.
DR ProDom; PD001947; CBD_3; 3.
DR SMART; SM00633; Glyco_10; 1.
DR PROSITE; PS00019; ACTININ_1; UNKNOWN_1.
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 1426 AA; 157543 MW; 29B3FDB85D09A863 CRC64;

Query Match 40.8%; Score 191.5; DB 2; Length 1426;
Best Local Similarity 43.5%; Pred. No. 5.6e-12;
Matches 37; Conservative 15; Mismatches 32; Indels 1; Gaps 1;

QY 4 GVKVQYKNDSAPGDNQIKPGLQVLNTGSSVDLSVTYVRYWFTRDGSGSTLVYNCDWAA 63
Db 118 GLKVLVYKNNETSASTSSIRPFKIVNGSSVDLSRVKIRYWTVDGDKPQSAV-CDWAQ 176
```



```
Db 413 GLKLVKNNETSASTSIRPFKIVNGSSVDLSRVKIRYWTVDGDKPQSAV-CDWAQ 471
Qy 64 MCGCNIRASFGSVNPTPTADTYLQ 88
Db 472 IGASNVTFNFVKLSGSGVGDYILE 496

RESULT 13
P96311
ID P96311 PRELIMINARY; PRT; 1711 AA.
AC P96311;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE 1,4-beta-glucanase (Fragment).
GN Name=ce1a;
OS Anaerococcus thermophilus.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Anaerococcus group;
OC Anaerococcus.
OX NCBI_TaxID=31899;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z-1320;
RX MEDLINE=98154434; PubMed=9493383;
RA Zverlov V., Mahr S., Riedel K., Bronnenmeier K.;
RT "Properties and gene structure of a bifunctional cellulolytic enzyme
RT (Cela) from the extreme thermophile Anaerococcus thermophilus with
RT separate glycosyl hydrolase family 9 and 48 catalytic domains.";
RL Microbiology 144:457-465(1998).
DR EMBL; Z86105; CAB06786.1; -.
DR PIR; T31337; T31337.
DR HSP; P37700; I987.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR005556; Glyco_hydro_48.
DR InterPro; IPR001701; Glyco_hydro_9.
DR InterPro; IPR008928; Glyco_trans_6hp.
DR InterPro; IPR002345; Lipocalin.
DR InterPro; IPR000209; Pept_S8_S53.
DR Pfam; PF00942; CBM_3; 3.
DR Pfam; PF02011; Glyco_hydro_48; 1.
DR Pfam; PF00759; Glyco_hydro_9; 1.
DR PRINTS; PR00844; GLHYDRLASE48.
DR ProDom; PD001947; CBD_3; 3.
DR ProDom; PD011903; Glyco_hydro_48; 1.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.
FT NON_TER
SQ SEQUENCE 1711 AA; 189979 MW; E3E987CEB9CD0C21 CRC64;

Query Match 40.7%; Score 191; DB 2; Length 1711;
Best Local Similarity 43.8%; Pred. No. 7.8e-12;
Matches 39; Conservative 18; Mismatches 30; Indels 2; Gaps 2;

Qy 1 VSGG-VKQVKNNDSPGDNQIKPGLQVNTGSSVDLSVTVYWFTRDGGSTLVNVC 59
Db 683 VAGQIKVLANKETNTTIRFLKLVKVTGSSIDLSRTIYWTVDGDKQAQSAIS- 741

Qy 60 DWAMGCGNIRASFGSVNPTPTADTYLQ 88
Db 742 DWAGIGASNVTFKFKVLSGSSVGDYILE 770

RESULT 14
Q9AQH0
ID Q9AQH0 PRELIMINARY; PRT; 996 AA.
AC Q9AQH0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
```

```
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Glycosyl hydrolase 5 (Fragment).
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
RT thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
DR EMBL; AF078038; AAK06388.1; -.
DR HSP; Q06851; INEC.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR002860; Glyco_hydro_BNR.
DR InterPro; IPR011040; Sialidase.
DR Pfam; PF02012; BNR; 9.
DR Pfam; PF00942; CBM_3; 1.
DR ProDom; PD001947; CBD_3; 1.
DR Hydrolase. 996 996
KW NON_TER
SQ SEQUENCE 996 AA; 108275 MW; 3C72B6ED22F3C614 CRC64;

Query Match 40.6%; Score 190.5; DB 2; Length 996;
Best Local Similarity 43.5%; Pred. No. 4.7e-12;
Matches 37; Conservative 15; Mismatches 32; Indels 1; Gaps 1;

Qy 4 GVKQVKNNDSPGDNQIKPGLQVNTGSSVDLSVTVYWFTRDGGSTLVNCDWAA 63
Db 844 GLKLVKNNETSASAGSIRPFKIVNGSSVDLSRVKIRYWTVDGDKPQSAV-CDWAQ 902

Qy 64 MCGCNIRASFGSVNPTPTADTYLQ 88
Db 903 IGASNVTFNFVKLSGSGVGDYILE 927

RESULT 15
Q52374
ID Q52374 PRELIMINARY; PRT; 1779 AA.
AC Q52374;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Family 10 xylanase (EC 3.2.1.8).
GN Name=xynC;
OS Caldicellulosiruptor sp. Rt69B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=70295;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rt69B.1;
RX MEDLINE=99283888; PubMed=10356996;
RA Morris D.D., Gibbs M.D., Ford M., Thomas J., Bergquist P.L.;
RT "Family 10 and 11 xylanase genes from Caldicellulosiruptor sp. strain
RT Rt69B.1.";
RL Extremophiles 3:103-111(1999).
CC -!- SIMILARITY: Belongs to cellulase family F (family 10 of glycosyl
CC hydrolases).
DR EMBL; AF036924; AAB95326.1; -.
DR PIR; T31085; T31085.
DR HSP; Q06851; INEC.
DR GO; GO:0030246; F:carbohydrate binding; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR006584; CBD_IV.
```



```

DR InterPro; IPR005084; CBM_6.
DR InterPro; IPR003305; CBM_CenC.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR008979; Gal_bind_like.
DR InterPro; IPR001000; Glyco_hydro_10.
DR InterPro; IPR006710; Glyco_hydro_43.
DR Pfam; PF00942; CBM_3; 3.
DR Pfam; PF02018; CBM_4_9; 2.
DR Pfam; PF03422; CBM_6; 1.
DR Pfam; PF00331; Glyco_hydro_10; 1.
DR Pfam; PF04616; Glyco_hydro_43; 1.
DR PRINTS; PR00134; GLHYDRLASE10.
DR ProDom; PD001947; CBD_3; 3.
DR SMART; SM00606; CBD_IV; 1.
DR SMART; SM00633; Glyco_10; 1.
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
KW Glycosidase; Hydrolase; Xylan degradation.
SQ SEQUENCE 1779 AA; 194304 MW; CE5269B6806B5CED CRC64;

Query Match 40.6%; Score 190.5; DB 2; Length 1779;
Best Local Similarity 43.5%; Pred. No. 9.3e-12;
Matches 37; Conservative 14; Mismatches 33; Indels 1; Gaps 1;

QY 4 GVKVOYKNDGAPGDNQIKPGLQLVNTGSSVDLSVTVYVWFTRDGSSSTLVVNCDWAA 63
Db 1113 GLKLYKNNETSASTGSRPWFKIVNGSSVDLSRVKIRYWTVDGDKPQSAV-CDWAQ 1171

QY 64 MCGGNIRASFGSVNPATPTADTYLQ 88
Db 1172 IGASNVTFNFVKLTSGVSGADYYLE 1196

```

Search completed: October 5, 2005, 07:56:09  
 Job time : 11.9413 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 07:49:28 ; Search time 3.381 Seconds  
(without alignments)  
1942.949 Million cell updates/sec

Title: US-09-917-376-5  
Perfect score: 469  
Sequence: 1 VSGGVKQVKNDSAPGNDQ.....IRASFGVNPATPTADTYLQ 88

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgm2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgm2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgm2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgm2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgm2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*  
6: /cgm2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	195.5	41.7	616	3	US-09-136-574A-47
2	194.5	41.5	1751	3	US-09-136-574A-44
3	191.5	40.8	1426	3	US-09-136-574A-43
4	172	36.7	700	2	US-07-862-588B-2
5	165.5	35.3	551	2	US-09-033-537A-1
6	165	35.2	167	5	PCT-US95-13813-9
7	165	35.2	476	4	US-09-339-159B-4
8	165	35.2	493	3	US-09-198-956-10
9	165	35.2	493	3	US-09-198-955A-12
10	165	35.2	493	3	US-09-694-531-12
11	165	35.2	493	4	US-09-670-141-10
12	165	35.2	493	4	US-10-072-152-12
13	155.5	33.2	1352	4	US-09-784-554B-2
14	150.5	32.1	1350	4	US-09-784-554B-4
15	120.5	25.7	531	2	US-07-862-588B-7
16	114	24.3	162	1	US-08-048-164A-2
17	114	24.3	162	1	US-08-460-462-2
18	114	24.3	162	1	US-08-460-457-2
19	114	24.3	162	1	US-08-460-458-2
20	114	24.3	162	2	US-08-460-455-2
21	114	24.3	162	2	US-08-330-394A-2
22	114	24.3	163	3	US-09-006-636-7
23	114	24.3	163	3	US-09-006-632-7
24	114	24.3	163	3	US-09-325-274-7
25	113	24.1	382	3	US-09-277-716-22
26	113	24.1	382	4	US-09-609-161B-22
27	112	23.9	154	2	US-08-330-394A-29
					Sequence 47, Appl
					Sequence 44, Appl
					Sequence 43, Appl
					Sequence 1, Appl
					Sequence 9, Appl
					Sequence 4, Appl
					Sequence 10, Appl
					Sequence 12, Appl
					Sequence 10, Appl
					Sequence 12, Appl
					Sequence 2, Appl
					Sequence 4, Appl
					Sequence 7, Appl
					Sequence 2, Appl
					Sequence 2, Appl
					Sequence 2, Appl
					Sequence 7, Appl
					Sequence 7, Appl
					Sequence 22, Appl
					Sequence 22, Appl
					Sequence 29, Appl

28	112	23.9	156	2	US-08-330-394A-22	Sequence 22, Appl
29	72.5	15.5	272	4	US-09-902-540-11056	Sequence 11056, A
30	65.5	14.0	1566	4	US-09-581-472B-2	Sequence 2, Appl
31	65	13.9	474	4	US-09-248-796A-25524	Sequence 25524, A
32	64	13.6	143	4	US-09-301-593-26	Sequence 26, Appl
33	64	13.6	428	3	US-09-118-319-5	Sequence 5, Appl
34	64	13.6	464	1	US-08-353-400-36	Sequence 36, Appl
35	64	13.6	472	4	US-09-301-593-30	Sequence 30, Appl
36	63.5	13.5	1785	3	US-09-341-587-3	Sequence 3, Appl
37	63	13.4	453	4	US-09-301-593-18	Sequence 18, Appl
38	63	13.4	472	4	US-09-301-593-43	Sequence 43, Appl
39	63	13.4	718	4	US-09-328-352-4640	Sequence 4640, Ap
40	63	13.4	1133	4	US-09-902-540-12243	Sequence 12243, A
41	63	13.4	1581	3	US-09-110-517-2	Sequence 2, Appl
42	62.5	13.3	288	3	US-09-423-439-38	Sequence 38, Appl
43	62.5	13.3	389	4	US-09-902-540-14347	Sequence 14347, A
44	62.5	13.3	445	1	US-08-353-400-33	Sequence 33, Appl
45	62.5	13.3	641	4	US-09-687-538B-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1  
US-09-136-574A-47  
; Sequence 47, Application US/09136574A  
; Patent No. 6294366  
; GENERAL INFORMATION:  
; APPLICANT: Farrington, Graham K.  
; Anderson, Paige  
; Gibbs, Moreland  
; Bergquist, Peter  
; Daniels, Roy  
; Morgan, Hugh W.  
; Williams, Diane P.  
; TITLE OF INVENTION: Compositions and Methods for  
; Treating Cellulose Containing  
; Cellulase Enzyme Compositions  
; NUMBER OF SEQUENCES: 49  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Howson and Howson  
; STREET: Spring House Corporate Center, P.O. Box 457  
; CITY: Spring House  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19477  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/136,574A  
; FILING DATE: 19-Aug-1998  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/932,571  
; FILING DATE: September 19, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bak, Mary E.  
; REGISTRATION NUMBER: 31,215  
; REFERENCE/DOCKET NUMBER: 1997US001/CIP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-540-9200  
; TELEFAX: 215-540-5818  
; TELEX: <Unknown>  
; INFORMATION FOR SEQ ID NO: 47:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 616 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 47:



US-09-136-574A-47

Query Match 41.7%; Score 195.5; DB 3; Length 616;  
Best Local Similarity 43.2%; Pred. No. 3.5e-13;  
Matches 38; Conservative 15; Mismatches 34; Indels 1; Gaps 1;

Qy 1 VSGGVKQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGDKPQSAV-CD 60  
Db 1 MGSQVLYKNNETSASTGSRPWFKIVNGSSVDLSRVKIRYWTVDGDKPQSAV-CD 59

Qy 61 WAAMGCNIRASFGSVNPTADTYLQ 88  
Db 60 WAQIGASNVTFNFVKLSGSGVADYYLE 87

RESULT 2  
US-09-136-574A-44  
; Sequence 44, Application US/09136574A  
; Patent No. 6294366  
; GENERAL INFORMATION:  
; APPLICANT: Farrington, Graham K.  
; Anderson, Paige  
; Gibbs, Moreland  
; Bergquist, Peter  
; Daniels, Roy  
; Morgan, Hugh W.  
; Williams, Diane P.  
; TITLE OF INVENTION: Compositions and Methods for  
; Treating Cellulose Containing Fabrics Using Truncated  
; Cellulase Enzyme Compositions  
; NUMBER OF SEQUENCES: 49  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Howson and Howson  
; STREET: Spring House Corporate Center, P.O. Box 457  
; CITY: Spring House  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19477  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/136,574A  
; FILING DATE: 19-Aug-1998  
; CLASSIFICATION: <Unknown>  
; PRIORITY INFORMATION:  
; PRIORITY NUMBER: US 08/932,571  
; FILING DATE: September 19, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bak, Mary E.  
; REGISTRATION NUMBER: 31,215  
; REFERENCE/DOCKET NUMBER: 1997US001/CIP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-540-9200  
; TELEFAX: 215-540-5818  
; TELEX: <Unknown>  
; INFORMATION FOR SEQ ID NO: 44:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1751 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 44:

Query Match 41.5%; Score 194.5; DB 3; Length 1751;  
Best Local Similarity 44.7%; Pred. No. 1.6e-12;  
Matches 38; Conservative 14; Mismatches 32; Indels 1; Gaps 1;

Qy 4 GVKVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGDKPQSAV-CD 63  
Db 4 GVKVQYKNNETSASTGSRPWFKIVNGSSVDLSRVKIRYWTVDGDKPQSAV-CD 63

Db 678 GVKVLYKNNETSASTGSRPWFKIVNGSSVDLSRVKIRYWTVDGDKPQSAV-CD 736

Qy 64 MGCNIRASFGSVNPTADTYLQ 88  
Db 737 IGASNVTFNFVKLSGSGVADYYLE 761

RESULT 3  
US-09-136-574A-43  
; Sequence 43, Application US/09136574A  
; Patent No. 6294366  
; GENERAL INFORMATION:  
; APPLICANT: Farrington, Graham K.  
; Anderson, Paige  
; Gibbs, Moreland  
; Bergquist, Peter  
; Daniels, Roy  
; Morgan, Hugh W.  
; Williams, Diane P.  
; TITLE OF INVENTION: Compositions and Methods for  
; Treating Cellulose Containing Fabrics Using Truncated  
; Cellulase Enzyme Compositions  
; NUMBER OF SEQUENCES: 49  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Howson and Howson  
; STREET: Spring House Corporate Center, P.O. Box 457  
; CITY: Spring House  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19477  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/136,574A  
; FILING DATE: 19-Aug-1998  
; CLASSIFICATION: <Unknown>  
; PRIORITY INFORMATION:  
; PRIORITY NUMBER: US 08/932,571  
; FILING DATE: September 19, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bak, Mary E.  
; REGISTRATION NUMBER: 31,215  
; REFERENCE/DOCKET NUMBER: 1997US001/CIP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-540-9200  
; TELEFAX: 215-540-5818  
; TELEX: <Unknown>  
; INFORMATION FOR SEQ ID NO: 43:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1426 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: No. 6294366e  
; SEQUENCE DESCRIPTION: SEQ ID NO: 43:

Query Match 40.8%; Score 191.5; DB 3; Length 1426;  
Best Local Similarity 43.5%; Pred. No. 2.7e-12;  
Matches 37; Conservative 15; Mismatches 32; Indels 1; Gaps 1;

Qy 4 GVKVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGDKPQSAV-CD 63  
Db 413 GLKVLKNNETSASTGSRPWFKIVNGSSVDLSRVKIRYWTVDGDKPQSAV-CD 471

Qy 64 MGCNIRASFGSVNPTADTYLQ 88  
Db 472 IGASNVTFNFVKLSGSGVADYYLE 496

RESULT 4



```
US-07-862-588B-2
; Sequence 2, Application US/07862588B
; Patent No. 5516796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linnaa
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862.588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-862-588B-2
Query Match 36.7%; Score 172; DB 2; Length 700;
Best Local Similarity 44.9%; Pred. No. 1.6e-10;
Matches 40; Conservative 18; Mismatches 27; Indels 4; Gaps 3;
QY 1 VSGGVKQVYKND-SAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGGSSTLVYNC 59
Db 549 VNSDLVQYKCDRRNATDQIKPHFNQKGTSPVLSLTLRYFTKD--SSAAMNGWI 607
QY 60 DWAMGCGNIRASFGSNPATPTADTYLQ 88
Db 608 DWAKLGGSNQISFGNHGA--DSDTYAE 634
RESULT 5
US-09-033-537A-1
; Sequence 1, Application US/09033537A
; Patent No. 5958083
; GENERAL INFORMATION:
; APPLICANT: Onishi, Masahiro
; APPLICANT: Fich, Merete
; APPLICANT: Toft, Annette Hanne
; APPLICANT: Sch lein, Martin
; TITLE OF INVENTION: Prevention Of Back-Staining
; TITLE OF INVENTION: In Stone Washing
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5958083o No. 5958083disk of No. 5958083th America, Inc.
```

```
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: U.S.A.
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/033,537A
FILING DATE: 02-MAR-1998
CLASSIFICATION: 008
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 0993/95
FILING DATE: 08-SEP-1995
APPLICATION NUMBER: PCT/DK96/00364
FILING DATE: 03-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Green, Reza
REGISTRATION NUMBER: 38,475
REFERENCE/DOCKET NUMBER: 4492.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 551 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-033-537A-1
Query Match 35.3%; Score 165.5; DB 2; Length 551;
Best Local Similarity 37.9%; Pred. No. 6e-10;
Matches 33; Conservative 22; Mismatches 29; Indels 3; Gaps 2;
QY 2 SGGVKVQYKND-SAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGGSSTLVYNC 61
Db 402 TGNLVVQYKVGDTSATDNQMKPSFNKNKNGTTPVNLGSLKLYFTKD-GTADMSASFDW 460
QY 62 AAMGCGNIRASFGSNPATPTADTYLQ 88
Db 461 AQIGASNVSAAP--ANFTGSDTYVE 485
RESULT 6
PCT-US95-13813-9
; Sequence 9, Application PC/TUS9513813
; GENERAL INFORMATION:
; APPLICANT: Yeda Research and Development Co. Ltd.
; APPLICANT: Ramot University Authority for Applied
; APPLICANT: Research and Industrial Development Ltd.
; APPLICANT: Technion Research and Development Foundation Ltd.
; APPLICANT: Bayer, Edward A.
; APPLICANT: Morag, Ely
; APPLICANT: Wilchek, Meir
; APPLICANT: Lamed, Raphael
; APPLICANT: Shoham, Yuval
; TITLE OF INVENTION: MODIFIED CELLULOSE-BINDING DOMAIN (CBD)
; TITLE OF INVENTION: PROTEINS AND USE THEREOF
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Broadway and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```











```

; SEQ ID NO 12
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Clostridium thermocellum
US-10-072-152-12

Query Match 35.2%; Score 165; DB 4; Length 493;
Best Local Similarity 36.1%; Pred. No. 6e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGVKVQYKNDSPAGNQIKPGLQLVNTGSSVDLSVTYVYFTRDGGSTLVNCD 60
Db 331 VSGNLKVFYNSPDSPTTNSINPQPKVTNTGSSAIDLSKLTLYYYTVDGQKQTFW-CD 389

QY 61 WAAM-----GCGNIRASFGSVNPATPTADTYLQ 88
Db 390 HAAIIGNSYNGITSNVKGFVVKMSSTNNADTYLE 426

RESULT 13
US-09-784-554B-2
; Sequence 2, Application US/09784554B
; Patent No. 6815192
; GENERAL INFORMATION:
; APPLICANT: Schnorr, Kirk
; APPLICANT: Jorgensen, Per Lina
; APPLICANT: Schulein, Martin
; TITLE OF INVENTION: FAMILY 44 XYLOGLUCANASES
; FILE REFERENCE: 10017.200-US
; CURRENT APPLICATION NUMBER: US/09/784,554B
; CURRENT FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1352
; TYPE: PRT
; ORGANISM: Paenibacillus polymyxa
US-09-784-554B-2

Query Match 33.2%; Score 155.5; DB 4; Length 1352;
Best Local Similarity 31.8%; Pred. No. 2.3e-08;
Matches 27; Conservative 21; Mismatches 36; Indels 1; Gaps 1;

QY 4 GVKVOYKNDSPAGNQIKPGLQLVNTGSSVDLSVTYVYFTRDGGSTLVNCDWAA 63
Db 1200 GLLQTRTADTKVNDNHLNPFQIVNKGITSPINELKIRIYYTIDGREQ-TFNCDYAT 1258

QY 64 MCGCNIRASFGSVNPATPTADTYLQ 88
Db 1259 LSCSKLNGKLVKMKAAATGADYYLE 1283

RESULT 14
US-09-784-554B-4
; Sequence 4, Application US/09784554B
; Patent No. 6815192
; GENERAL INFORMATION:
; APPLICANT: Schnorr, Kirk
; APPLICANT: Jorgensen, Per Lina
; APPLICANT: Schulein, Martin
; TITLE OF INVENTION: FAMILY 44 XYLOGLUCANASES
; FILE REFERENCE: 10017.200-US
; CURRENT APPLICATION NUMBER: US/09/784,554B
; CURRENT FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 1350
; TYPE: PRT
; ORGANISM: Paenibacillus polymyxa
US-09-784-554B-4

Query Match 32.1%; Score 150.5; DB 4; Length 1350;

```

```

Best Local Similarity 31.8%; Pred. No. 8.1e-08;
Matches 27; Conservative 21; Mismatches 36; Indels 1; Gaps 1;

QY 4 GVKVOYKNDSPAGNQIKPGLQLVNTGSSVDLSVTYVYFTRDGGSTLVNCDWAA 63
Db 1199 GVLQTRTADTKVNDNHLNPFQIVNKGITSPINELKIRIYYTIDGREQ-TFNCDYAV 1257

QY 64 MCGCNIRASFGSVNPATPTADTYLQ 88
Db 1258 LSCSKLNGKLVKMKAAATGADYYLE 1282

RESULT 15
US-07-862-588B-7
; Sequence 7, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linaa
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5916796o No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 531 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-862-588B-7

Query Match 25.7%; Score 120.5; DB 2; Length 531;
Best Local Similarity 32.9%; Pred. No. 5e-05;
Matches 28; Conservative 18; Mismatches 36; Indels 3; Gaps 2;

QY 2 SGGVKVOYKNDSPAGNQIKPGLQLVNTGSSVDLSVTYVYFTRDGGSTLVNCDW 61
Db 432 TGNLVVQYKVGDTSATDNQKPSFNKKNNGTTPVNL-SGLKXXXXKD-GPADMSCSIDW 490

QY 62 AAMGCGNIRASFGSVNPATPTADTY 86
Db 491 AQIGRTNVLAF--ANFTGSNTDITY 513

Search completed: October 5, 2005, 08:04:32

```



us-09-917-376-5.ra1

Thu Oct 6 11:20:49 2005

Job time : 4.381 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 07:53:23 ; Search time 11.1761 Seconds  
(without alignments)  
3288.602 Million cell updates/sec

Title: US-09-917-376-5  
Perfect score: 469  
Sequence: 1 VSGGVKQYKKNDSAPGDNQ.....IRASFGSVNPATPTADTYLQ 88

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1846076 seqs, 415116000 residues  
Total number of hits satisfying chosen parameters: 1846076

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA.\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US10E\_PUBCOMB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/1/pubpaa/US11A\_PUBCOMB.pep.\*  
20: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*  
21: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
22: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	469	100.0	88	11	US-09-917-376-5
2	469	100.0	88	14	US-10-155-400-5
3	469	100.0	89	11	US-09-917-376-4
4	469	100.0	89	14	US-10-155-400-4
5	469	100.0	154	10	US-09-917-378-4
6	469	100.0	762	10	US-09-917-378-1
7	466	99.4	150	10	US-09-917-384-5
8	466	99.4	150	10	US-09-917-383-5
9	466	99.4	1043	10	US-09-917-384-6
10	466	99.4	1043	10	US-09-917-383-6
11	466	99.4	1228	10	US-09-917-384-1

12	466	99.4	1228	10	US-09-917-383-1	Sequence 1, Appli
13	462	98.5	957	11	US-09-917-376-1	Sequence 1, Appli
14	462	98.5	957	14	US-10-155-400-1	Sequence 1, Appli
15	257	54.8	741	14	US-10-156-761-8100	Sequence 8100, Ap
16	190	40.5	508	15	US-10-369-493-23151	Sequence 23151, A
17	188	40.1	477	16	US-10-466-208-12	Sequence 12, Appli
18	188	40.1	496	16	US-10-466-208-8	Sequence 8, Appli
19	188	40.1	677	16	US-10-433-577-35	Sequence 35, Appli
20	180	38.4	1621	14	US-10-185-990-10	Sequence 10, Appli
21	165	35.2	167	17	US-10-933-404-4	Sequence 4, Appli
22	165	35.2	476	15	US-10-372-054-4	Sequence 4, Appli
23	165	35.2	493	13	US-10-072-152-12	Sequence 12, Appli
24	165	35.2	493	15	US-10-655-433-12	Sequence 12, Appli
25	165	35.2	599	10	US-09-955-555A-29	Sequence 29, Appli
26	155.5	33.2	1352	10	US-09-784-554B-2	Sequence 2, Appli
27	155.5	33.2	1352	16	US-10-896-555-2	Sequence 2, Appli
28	150.5	32.1	1350	10	US-09-784-554B-4	Sequence 4, Appli
29	150.5	32.1	1350	16	US-10-896-555-4	Sequence 4, Appli
30	130.5	27.8	1483	15	US-10-282-122A-51483	Sequence 51483, A
31	114	24.3	256	14	US-10-261-446-6	Sequence 6, Appli
32	114	24.3	256	15	US-10-261-445B-6	Sequence 6, Appli
33	114	24.3	256	17	US-10-782-234-6	Sequence 6, Appli
34	113	24.1	382	10	US-09-808-898-22	Sequence 22, Appli
35	105.5	22.5	163	15	US-10-460-524-2	Sequence 2, Appli
36	78.5	16.7	1049	15	US-10-282-122A-49900	Sequence 49900, A
37	75.5	16.1	346	16	US-10-437-963-179234	Sequence 223, App
38	75.5	16.1	618	15	US-10-211-462-223	Sequence 4039, Ap
39	75.5	16.1	618	16	US-10-723-860-4039	Sequence 5713, Ap
40	75.5	16.1	618	18	US-10-756-149-5713	Sequence 53, Appli
41	75.5	16.1	986	9	US-09-747-835A-53	Sequence 101, App
42	75.5	16.1	986	14	US-10-120-604-101	Sequence 406, App
43	75.5	16.1	986	14	US-10-225-567A-406	Sequence 3, Appli
44	75.5	16.1	986	15	US-10-398-458-3	Sequence 53, Appli
45	75.5	16.1	986	15	US-10-312-312-53	

ALIGNMENTS

RESULT 1  
US-09-917-376-5  
; Sequence 5, Application US/09917376  
; Publication No. US20040038334A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: 40197.4US01  
; CURRENT APPLICATION NUMBER: US/09/917,376  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 88  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURES:  
; OTHER INFORMATION: Carbohydrate binding domain  
US-09-917-376-5

Query Match	100.0%	Score	469	DB	11	Length	88
Best Local Similarity	100.0%	Pred. No.	3.6e-48				
Matches	88	Conservative	0	Mismatches	0	Indels	0
Qy	1	VSGGVKQYKKNDSAPGDNQIKPGLQIVNTGSSVDLSTVTYVWFTDGGSSITLVNCD	60				
Db	1	VSGGVKQYKKNDSAPGDNQIKPGLQIVNTGSSVDLSTVTYVWFTDGGSSITLVNCD	60				
Qy	61	WAAMGCNTRASFSGSVNPATPTADTYLQ	88				



```

Db 61 WAAMCGCNIRASFGSVNPATPTADTYLQ 88
RESULT 2
US-10-155-400-5
; Sequence 5, Application US/10155400
; Publication No. US20030108988A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOYTICUS
; FILE REFERENCE: NREL 01-36A
; CURRENT APPLICATION NUMBER: US/10/155,400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 88
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; OTHER INFORMATION: Carbohydrate binding domain
US-10-155-400-5
Query Match 100.0%; Score 469; DB 14; Length 88;
Best Local Similarity 100.0%; Pred. No. 3.6e-48;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVKYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLYVNC 60
Db 1 VSGGVKQVKYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLYVNC 60

QY 61 WAAMCGCNIRASFGSVNPATPTADTYLQ 88
Db 61 WAAMCGCNIRASFGSVNPATPTADTYLQ 88
RESULT 3
US-09-917-376-4
; Sequence 4, Application US/09917376
; Publication No. US20040038334A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOYTICUS
; FILE REFERENCE: 40197.4US01
; CURRENT APPLICATION NUMBER: US/09/917,376
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 89
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (89)
; OTHER INFORMATION: Any amino acid
; OTHER INFORMATION: Carbohydrate binding domain
US-09-917-376-4
Query Match 100.0%; Score 469; DB 11; Length 89;
Best Local Similarity 100.0%; Pred. No. 3.6e-48;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVKYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLYVNC 60
Db 1 VSGGVKQVKYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLYVNC 60

QY 61 WAAMCGCNIRASFGSVNPATPTADTYLQ 88
Db 61 WAAMCGCNIRASFGSVNPATPTADTYLQ 88
RESULT 4
US-10-155-400-4
; Sequence 4, Application US/10155400
; Publication No. US20030108988A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOYTICUS
; FILE REFERENCE: NREL 01-36A
; CURRENT APPLICATION NUMBER: US/10/155,400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 89
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (89)
; OTHER INFORMATION: Any amino acid
; OTHER INFORMATION: Carbohydrate binding domain
US-10-155-400-4
Query Match 100.0%; Score 469; DB 14; Length 89;
Best Local Similarity 100.0%; Pred. No. 3.6e-48;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVKYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLYVNC 60
Db 1 VSGGVKQVKYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLYVNC 60

QY 61 WAAMCGCNIRASFGSVNPATPTADTYLQ 88
Db 61 WAAMCGCNIRASFGSVNPATPTADTYLQ 88
RESULT 5
US-09-917-378-4
; Sequence 4, Application US/09917378
; Publication No. US20030119093A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT MANNANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOYTICUS
; FILE REFERENCE: 40197.7US01
; CURRENT APPLICATION NUMBER: US/09/917,378
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Carbohydrate
; OTHER INFORMATION: binding domain
US-09-917-378-4

```



```

; OTHER INFORMATION: Guxa
US-09-917-384-5
Query Match          99.4%; Score 466; DB 10; Length 150;
Best Local Similarity 98.9%; Pred. No. 1.6e-47;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGGSTLVYVNC 60
    |||||
Db 1 VSGGLKVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGGSTLVYVNC 60
    |||||

QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
    |||||
Db 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
    |||||

RESULT 8
US-09-917-383-5
; Sequence 5, Application US/09917383
; Publication No. US20030104522A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,383
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-383-5
Query Match          99.4%; Score 466; DB 10; Length 150;
Best Local Similarity 98.9%; Pred. No. 1.6e-47;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGGSTLVYVNC 60
    |||||
Db 1 VSGGLKVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGGSTLVYVNC 60
    |||||

QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
    |||||
Db 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
    |||||

RESULT 9
US-09-917-384-6
; Sequence 6, Application US/09917384
; Publication No. US20030096342A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,384
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
```

```

Query Match          100.0%; Score 469; DB 10; Length 154;
Best Local Similarity 100.0%; Pred. No. 7e-48;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGGSTLVYVNC 60
    |||||
Db 1 VSGGVKQVQKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGGSTLVYVNC 60
    |||||

QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
    |||||
Db 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
    |||||

RESULT 6
US-09-917-378-1
; Sequence 1, Application US/09917378
; Publication No. US20030119093A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT MANNANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40197.7US01
; CURRENT APPLICATION NUMBER: US/09/917,378
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 762
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
US-09-917-378-1
Query Match          100.0%; Score 469; DB 10; Length 762;
Best Local Similarity 100.0%; Pred. No. 4.9e-47;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGGSTLVYVNC 60
    |||||
Db 455 VSGGVKQVQKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGGSTLVYVNC 514
    |||||

QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
    |||||
Db 515 WAAMGCGNIRASFGSVNPATPTADTYLQ 542
    |||||

RESULT 7
US-09-917-384-5
; Sequence 5, Application US/09917384
; Publication No. US20030096342A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,384
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
```



```
; LENGTH: 1043
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-384-6

Query Match          99.4%; Score 466; DB 10; Length 1043;
Best Local Similarity 98.9%; Pred. No. 1.6e-46;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLVNCD 60
   |||:|||||
Db 477 VSGGLKVQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLVNCD 536
   |||:|||||

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
   |||:|||||
Db 537 WAAMCGNIRASFGSVNPATPTADTYLQ 564
   |||:|||||

RESULT 10
US-09-917-383-6
; Sequence 6, Application US/09917383
; Publication No. US20030104522A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOXYLICUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,383
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1043
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-383-6

Query Match          99.4%; Score 466; DB 10; Length 1043;
Best Local Similarity 98.9%; Pred. No. 1.6e-46;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLVNCD 60
   |||:|||||
Db 477 VSGGLKVQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLVNCD 536
   |||:|||||

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
   |||:|||||
Db 537 WAAMCGNIRASFGSVNPATPTADTYLQ 564
   |||:|||||

RESULT 11
US-09-917-384-1
; Sequence 1, Application US/09917384
; Publication No. US20030096342A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOXYLICUS
; FILE REFERENCE: 40170.6US01

; LENGTH: 1043
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-384-1

Query Match          99.4%; Score 466; DB 10; Length 1043;
Best Local Similarity 98.9%; Pred. No. 1.6e-46;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLVNCD 60
   |||:|||||
Db 477 VSGGLKVQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLVNCD 536
   |||:|||||

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
   |||:|||||
Db 537 WAAMCGNIRASFGSVNPATPTADTYLQ 564
   |||:|||||

RESULT 12
US-09-917-383-1
; Sequence 1, Application US/09917383
; Publication No. US20030104522A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOXYLICUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,383
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1228
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-383-1

Query Match          99.4%; Score 466; DB 10; Length 1228;
Best Local Similarity 98.9%; Pred. No. 2e-46;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLVNCD 60
   |||:|||||
Db 584 VSGGLKVQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLVNCD 643
   |||:|||||

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
   |||:|||||
Db 644 WAAMCGNIRASFGSVNPATPTADTYLQ 671
   |||:|||||

RESULT 13
US-09-917-376-1
; Sequence 1, Application US/09917376
; Publication No. US20040038334A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
```



```
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40197.4US01
; CURRENT APPLICATION NUMBER: US/09/917.376
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 957
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (957)
; OTHER INFORMATION: Any amino acid
US-09-917-376-1

Query Match      98.5%; Score 462; DB 11; Length 957;
Best Local Similarity 97.7%; Pred. No. 4.4e-46;
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVYVNC 60
Db 869 VSGGVKQYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVYVNC 928

QY 61 WAAMCGCNIRASFGSVNPATPTADTYLQ 88
Db 929 WAAIGCGNIRASFGSVNPATPTADTYLQ 956

RESULT 14
US-10-155-400-1
; Sequence 1, Application US/10155400
; Publication No. US20030108988A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: NREL 01-36A
; CURRENT APPLICATION NUMBER: US/10/155,400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 957
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (957)
; OTHER INFORMATION: Any amino acid
US-10-155-400-1

Query Match      98.5%; Score 462; DB 14; Length 957;
Best Local Similarity 97.7%; Pred. No. 4.4e-46;
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVYVNC 60
Db 869 VSGGVKQYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVYVNC 928

QY 61 WAAMCGCNIRASFGSVNPATPTADTYLQ 88
Db 929 WAAIGCGNIRASFGSVNPATPTADTYLQ 956

RESULT 15
US-10-156-761-8100
; Sequence 8100, Application US/10156761
```

```
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8100
; LENGTH: 741
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8100

Query Match      54.8%; Score 257; DB 14; Length 741;
Best Local Similarity 56.3%; Pred. No. 1.2e-21;
Matches 49; Conservative 13; Mismatches 25; Indels 0; Gaps 0;

QY 2 SGGVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVYVNC 61
Db 591 SGGKLVLYKNDSSATDNNAIRPGLRIVNTGSGSLDLSKVTRYFTTRDGSPTVNAWCDY 650

QY 62 AAAMCGCNIRASFGSVNPATPTADTYLQ 88
Db 651 AAAGCNSVSLKVVPVLTTPVPGADAYLE 677

Search completed: October 5, 2005, 08:08:32
Job time : 11.1761 secs
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 5, 2005, 07:09:01 ; Search time 1158.46 Seconds  
(without alignments)  
3680.789 Million cell updates/sec

Title: US-09-917-376-5  
Perfect score: 469  
Sequence: 1 VSGGVKQYKXNDAPGDQ.....IRASFGSVNPATPTDYLQ 88

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-O=/cgn2.1/USPTO.spool/US09917376/runat\_04102005.164328.26897/app.query.fasta.1.2524  
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09917376 @CNC 1.1.15287 @runat\_04102005.164328.26897 -NCPU=3  
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.\*

1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sv.\*  
13: gb.un.\*  
14: gb.vi.\*

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	469	100.0	2289	6 AX700050	Sequence
2	466	99.4	3687	6 AX700036	Sequence
3	462	98.5	2869	6 AX700058	Sequence
4	436	93.0	3365	6 AX700025	Sequence

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

5	257	54.8	299175	1	AP005023	AP005023 Streptomy
6	228.5	48.7	1957	1	AY298814	AY298814 Thermobif
7	225.5	48.1	4567	1	AF163837	AF163837 Caldicell
8	214.5	45.7	3237	1	AF200304	AF200304 Caldicell
9	209.5	44.7	3509	1	PAE488933	PAE488933 Paenibaci
10	204.5	43.6	4161	1	BSF133614	BSF133614 Bacillus
11	195.5	41.7	2029	6	E35142	E35142 Truncated c
12	195.5	41.7	2029	6	E35143	E35143 Truncated c
13	194.5	41.5	6005	1	AF078038S5	AF078038 Caldicell
14	194.5	41.5	6416	6	E35100	E35100 Truncated c
15	192.5	41.0	4743	1	AB008029	AB008029 Thermophi
16	192.5	41.0	11707	1	AF078737	AF078737 Caldicell
17	192.5	41.0	11707	6	E35099	E35099 Truncated c
18	191.5	40.8	787	1	AF078038S3	AF078038 Caldicell
19	191	40.7	5513	1	AT286105	AT286105 A.thermophi
20	190.5	40.6	3262	1	AF078038S1	AF078038 Caldicell
21	190.5	40.6	5437	1	CASR69XN2	AF036924 Caldicell
22	190	40.5	1523	1	AF355629	AF355629 Bacillus
23	190	40.5	1800	1	BSEGLSG	Z29076 Bacillus su
24	190	40.5	1920	1	BACGLUB	M16185 B.subtilis
25	190	40.5	2314	1	BS14GLUC	X67044 B.subtilis
26	190	40.5	2435	1	BSBGLUC2	X04689 Bacillus su
27	190	40.5	2589	1	AY044252	AY044252 Bacillus
28	190	40.5	26170	1	BC170DEGR	Z73234 B.subtilis
29	190	40.5	122883	1	BAM575417	AJ576102 Bacillus
30	190	40.5	122883	1	BSUB0010	Z99113 Bacillus su
31	190	40.5	207829	1	BACCMCASE	D01057 B.subtilis
32	188	40.1	1928	1	BACCMCASE	AY183475 Bacillus
33	187	39.9	1593	1	AY183475	AF045482 Bacillus
34	187	39.9	2084	1	AF045482	L01257 Caldocellum
35	186.5	39.8	4977	1	CDQWANA	L32742 Caldocellum
36	186.5	39.8	5439	1	CDCELA	M28332 B.subtilis
37	186	39.7	2175	1	BACCELD	E05425 DNA sequenc
38	186	39.7	2175	6	E05425	AB016164 Bacillus
39	186	39.7	2587	1	AB016164	A28170 B.lautus st
40	182.5	38.9	2977	6	A28170	X13602 Caldocellum
41	182.5	38.9	4241	1	CSCBLB	U16308 Caldocellum
42	182.5	38.9	5284	1	CSU16308	M76598 Bacillus la
43	180.5	38.5	2831	1	BACCELA	U27084 Bacillus sp
44	180	38.4	1553	1	BSU27084	AF363635 Bacillus
45	177	37.7	2152	1	AF363635	

#### ALIGNMENTS

RESULT 1	AX700050	Sequence 2 from Patent WO03012110.	2289 bp	DNA	linear	PAT 03-APR-2003
LOCUS	AX700050	Sequence 2 from Patent WO03012110.				
DEFINITION	AX700050					
ACCESSION	AX700050					
VERSION	AX700050.1	GI:29536020				
KEYWORDS						
SOURCE						
ORGANISM						
		Acidothermus cellulolyticus				
		Acidothermus cellulolyticus				
		Bacteria: Actinobacteria; Actinobacteridae; Actinomycetales;				
		Frankineae; Acidothermaceae; Acidothermus.				
REFERENCE	1					
AUTHORS		Ding, S.Y., Adney, W.S., Vinzant, T.B. and Himmel, M.E.				
TITLE		Thermal tolerant mannanase from Acidothermus cellulolyticus				
JOURNAL		Patent: WO 03012110-A 2 13-FEB-2003;				
		Midwest Research Institute (US)				
FEATURES						
source		Location/Qualifiers				
		1..2289				
		/organism="Acidothermus cellulolyticus"				
		/mol_type="unassigned DNA"				
		/db_xref="taxon:28049"				

#### ORIGIN

Alignment Scores:					
Pred. No.:	3.59e-48	Length:	2289		
Score:	469.00	Matches:	88		
Percent Similarity:	100.00%	Conservative:	0		
Best Local Similarity:	100.00%	Mismatches:	0		











CDS	YGIASLSYLSGVGNDRARYLAEQALHTLLQWRTTETPANGRPASS" 2694..3098 /note="SAV449" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="BAC68159.1" /db_xref="GI:29604088" /translation="MNYPAPASGIEBQRFQWGLDEYLMWCSDDIIADLAERGLRQLPLV HGGDYTCAGVFWVHLDGLFAYLTHPTTYQTDHLGLHVERLEIFDVSDDDDDPDLNAX IQELHRRNELNPQIAGMLADVPYTGSGVDD" 3127..4029 /note="SAV450" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="BAC68160.1" /db_xref="GI:29604099" /translation="MAKHVIIGAGSIGTNVARLLAEGGESVRIIVTRSGSGPEHPLIVDR VAADASPSRLTSLSGAGVITYHCANPPSYTWERLLPPLQTAATIAAKANDAVIALT GSLVAYGPQPGRMNEHTPMAATGHKGLRRRMWEOALAGGIRTVVRGSDYIGKDN GIYSLIKSALQKGAWITGHLDMPHPTFTENGMAOALITLASERAWGRPHVPSA PAVIRELARIYAVNAGRPPIKLQIPRSVTRTAGLIVPIAREMAEMDYQMYAPFHMD ATETTFGLTATDLDTAVRNEVG" complement(4130..4498) /gene="bldg" complement(4130..4498) /gene="bldg" /note="SAV451" /codon_start=1 /transl_table=11 /product="putative anti-anti-sigma factor" /protein_id="BAC68161.1" /db_xref="GI:29604090" /translation="MTDAFHIDVPHDDAALAVIALSGSFDIAAAPVARARALELIANG HPDLVADLSGYTFCDSGLGALVIGWRCADKADGSLTLAAIPDLRLSLVSTGMDAFL PAYSDAALAAARQGNRTTA" 4635..5156 /note="SAV452" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="BAC68162.1" /db_xref="GI:29604091" /translation="MFAVKESRVVGEIAATPTGEALELVLHGESLRGVRSDDA TARPDGLDFALPARGWHTLIVRPSVEGEEAGGLFHPGAEPGESGPARLCAWRE SAPRTIDEDVEVHTLRSLDILGAPHLRPAAPRTHGRRRCPLVHGVRPRLLAAV LVHGTAAQFCTPR" 5183..5665 /note="SAV453" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="BAC68163.1" /db_xref="GI:29604092" /translation="MVTAVRARAGSAWQACGRBSSQHSSEGVREMDKEFLRTV AERGLSQEADLTWATLETLARLSPEARDLITELPEGLAEAVRGTDRIERFG HDSVRRVAERTMLKEEADRGVRTVLAVLEAISEKEFNLDMSQLGTFDYSQALESAG " 5810..6964 /note="SAV454" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="BAC68164.1" /db_xref="GI:29604093" /translation="WIMTADVPTSSGRSVRVALVADPGLPSEIARALAPRLPHLRR AVGARVRQVDTWAPLSVSEQADVSDIAEAVRSHLDGAENDVAIPLTDFPRRLRYP ISVEVDTRAAALISLPALGVRLLRRVRQAVDVVRELVAEHTPPRPEVIGRPPE FVPDTRPGEPQVRVGLRGHTLRVAGMYANRPRLFGSLSRALAGVPATATVLF INSATNATLQAGVAQHVIAGSAVALTLWIIVDHQLWERGTHLPARHHPLYPNLV TLVITALAVVLLAALFATLVAVSFLILDASVLRKFIGRVGVGDFLFLAWFITAIM IGGAFGTSLEGDEKVDAAVGRQRORQMLRQMPRPERGERGPDITDG"
CDS	complement(6986..7498) /note="SAV455" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="BAC68165.1" /db_xref="GI:29604094" /translation="MNPYPAPASGIEBQRFQWGLDEYLMWCSDDIIADLAERGLRQLPLV ATVARLEAGERVLAIVGPQRTWTRALGDRYMHQNAVAFGSPAFVLGSMASWATSALG NSARRAANAQAPRCGARTPGLHLVLDVLDVHGHTMLRWSMHLGTSSSTRPSSALRALL ELVDLHVIAL" 7548..8189 /note="SAV456" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="BAC68166.1" /db_xref="GI:29604095" /translation="MATLLRRVCHEGVRIIAVPAGFVLSRRPWTASVYRCVAVGQPS EDRREVEGSRTPVASASSRLRPCCRLLRVPVRRAVPSQAVRSRFPRTGTRPLGDLV VYORHRLAHARVHVIIIDVHPGLGRGAARPVRIATHSLGGVIAVDMATANVPLMTES LVTPGSQATFFHQPVPGDLLQCMIAFRMRKHQEAETGALPPETGFIHAGQGW"
CDS	complement(6986..7498) /note="SAV455" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="BAC68165.1" /db_xref="GI:29604094" /translation="MNPYPAPASGIEBQRFQWGLDEYLMWCSDDIIADLAERGLRQLPLV ATVARLEAGERVLAIVGPQRTWTRALGDRYMHQNAVAFGSPAFVLGSMASWATSALG NSARRAANAQAPRCGARTPGLHLVLDVLDVHGHTMLRWSMHLGTSSSTRPSSALRALL ELVDLHVIAL" 7548..8189 /note="SAV456" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="BAC68166.1" /db_xref="GI:29604095" /translation="MATLLRRVCHEGVRIIAVPAGFVLSRRPWTASVYRCVAVGQPS EDRREVEGSRTPVASASSRLRPCCRLLRVPVRRAVPSQAVRSRFPRTGTRPLGDLV VYORHRLAHARVHVIIIDVHPGLGRGAARPVRIATHSLGGVIAVDMATANVPLMTES LVTPGSQATFFHQPVPGDLLQCMIAFRMRKHQEAETGALPPETGFIHAGQGW"
gene	Alignment Scores: Pred. No.: 4,31e-19 Length: 299175 Score: 257.00 Matches: 49 Percent Similarity: 71.26% Conservative: 13 Best Local Similarity: 56.32% Mismatches: 25 Query Match: 54.80% Indels: 0 DB: 1 Gaps: 0
CDS	US-09-917-376-5 (1-88) x AP005023 (1-299175) Qy 2 SerGlyGlyValIleValGlnTyrIlyAsnAsnAspSerAlaProGlyAspGlnIle 21 Db 105341 TCGGCGCGCTTCCAGGTCTCTACAGAACCAACGACTCTCTCGGCCACCGACATC 105400 Qy 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41 Db 105401 CGGCCAGCGCTTCGATGTCACACCGGACGCGCTCCCTCGACTGTCCAAGGTACG 105460 Qy 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61 Db 105461 GCCCGCTACTACTTCAACCGGACAGCGCTCGCCACCGTGAACGCGTGGTGGACTAC 105520 Qy 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81 Db 105521 GCGGCGCTCGGTGTTCACACGCTCAGAGGTCTGATCCCTGACCCGCGCGCG 105580 Qy 82 ThrAlaAspThrTyrLeuGln 88 Db 105581 GGAGCCGACGCTACCTCGAA 105601
LOCUS	AY298814 1957 bp DNA linear BCT 23-JUN-2003
DEFINITION	Thermobifida fusca strain TMS1 endoglucanase (cel5B) gene, complete cds.
ACCESSION	AY298814
VERSION	AY298814.1 GI:31745732
KEYWORDS	Thermobifida fusca
SOURCE	Thermobifida fusca
ORGANISM	Thermobifida fusca Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptosporangineae; Nocardiopsaceae; Thermobifida.
REFERENCE	1 (bases 1 to 1957) Posta, K., Beki, E., Kukolya, J. and Hornok, L. Phylogenetic relationships of T. cel5B, a new endoglucanase encoding gene from thermobifida fusca
AUTHORS	Unpublished
TITLE	2 (bases 1 to 1957) Posta, K., Beki, E., Kukolya, J. and Hornok, L. Direct Submission
JOURNAL	
REFERENCE	
AUTHORS	
TITLE	



JOURNAL Submitted (14-MAY-2003) Agricultural Biotechnology and Microbiology, Szent Istvan University Godollo, Pater Karoly str.1., Godollo 2103, Hungary

## FEATURES

source Location/Qualifiers

1. .1957 /organism="Thermobifida fusca"

/mol\_type="genomic DNA"

/strain="TW51"

/db\_xref="taxon:2021"

26. .39

/note="putative"

/bound\_moiety="CelR"

107. .1957

/gene="cel5B"

107. .1957

/gene="cel5B"

/codon\_start=1

/transl\_table=11

/product="endoglucanase"

/protein\_id="AAP56348.1"

/db\_xref="GI:31745733"

/translation="MTPLTRLRAGAAAIATGASALIPLTSSPAAAGTADWLHTDGN  
RIVDSAGNEVLTCANWFGNTSERMFGHWAANIEDITSAMERGINMVRPISTOL  
LLEWNGCAGSCVNEVYNPLAGMNTLEVDFWLQCEEYGLKVLMDVSAEADNSG  
HYIPWYKGDITTEFTANERWTERYKNDTIIVADIKNPHGKANETPRKWDGST  
DIDNFKVCTEAGKRIALINPMILICGIBIIPDQGDSSSTQGRDITSTWNGNLR  
GADHPVDLGAHQDQVYPHYDGYSPVEQFEGEWMNRQTLTEDVWRPNWLYIHEDD  
IAPLLIGMGWGLDQGNKWTALRSIIIDEKMHFTWALPNPSGDTGLLNDWTIT  
WDEAKYALFKALMDQANGKFGVLDHVDPLGCVGSTGVSLNOYVGGGPGQPTEPT  
EPTEPTPEPTPEPTPEPTPEPTPEPTPEPTPEPTPEPTPEPTPEPTPEPTPEPT  
VEIHYFTFEGGTLQFTCDWAQVGCANVNASFTSLAPGADTSLVLTSLSLAPGAS  
TELQQRHITANWANFDESDYSRGNTDWEIUSEVITVLGGTLVWGTPPA"

misc\_feature

236. .1258

/gene="cel5B"

/note="Region: catalytic domain; GH5 family"

1532. .1765

/gene="cel5B"

/note="Region: cellulose binding domain; CBM-3"

## ORIGIN

Alignment Scores:

Pred. No.: 2,788-18 Length: 1957  
Score: 228.50 Matches: 45  
Percent Similarity: 70.93% Conservative: 16  
Best Local Similarity: 52.33% Mismatches: 22  
Query Match: 48.72% Indels: 3  
DB: 1 Gaps: 2

US-09-917-376-5 (1-88) x AY298814 (1-1957)

QY 2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21  
Db 1517 ACGGTGGCTCGAGTCTACTACCGCAACACAGCTTTCGGCCGACGACGACGATC 1576  
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41  
Db 1577 GCGCGGGCTCGGCTGGTCAACACCGGACGACGACGCTCGCTCGCTCGGAA 1636  
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61  
Db 1637 ATCCACTACTACTTACCAACAGCGCCGCGGCGGCGGCGGCGGCGGCGGCGG 1693  
QY 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81  
Db 1694 GCTCAAGTGGGCTGGCCCAACGCTCAACGCGTCTCTTCACTCGCTCGCTG 1747  
QY 82 ThrAlaAspThrTyrLeu 87  
Db 1748 GCGCGCGACACCTCCCTG 1765

## RESULT 7

AF163837 4567 bp DNA linear BCT 08-FEB-2000  
LOCUS

## DEFINITION

Caldibacillus cellulovorans multidomain beta-1,4-mannanase precursor (mana) gene, complete cds; and unknown genes.

## ACCESSION

AF163837

## VERSION

AF163837.1 GI:6651325

## KEYWORDS

Caldibacillus cellulovorans

## SOURCE

Caldibacillus cellulovorans

## ORGANISM

Bacteria; Firmicutes; Bacillales; Alicyclobacillaceae;

## REFERENCE

1 (bases 1 to 4567)

## AUTHORS

Sunna,A., Gibbs,M.D., Chin,C.W., Nelson,P.J. and Bergquist,P.L.

## TITLE

A gene encoding a novel multidomain beta-1,4-mannanase from

## JOURNAL

Caldibacillus cellulovorans and action of the recombinant enzyme on

## PUBLISHED

kraft pulp

## MEDLINE

Appl. Environ. Microbiol. 66 (2), 664-670 (2000)

## PUBMED

20120520

## REFERENCE

10653733

## AUTHORS

2 (bases 1 to 4567)

## TITLE

Sunna,A., Gibbs,M.D. and Bergquist,P.L.

## JOURNAL

Submitted (29-JUN-1999) Biological Sciences, Macquarie University,

## PUBLISHED

North Ryde, New South Wales 2109, Australia

## FEATURES

Location/Qualifiers

## source

1. .4567

/organism="Caldibacillus cellulovorans"

/mol\_type="genomic DNA"

/db\_xref="taxon:74586"

<1. .513

/note="orf1"

/codon\_start=1

/transl\_table=11

/product="unknown"

/protein\_id="AAF22273.1"

/db\_xref="GI:6651326"

/translation="TPTPTPSAAPTPFSAGSLVQVRAADTNAGDNQLKPHFRI

VNRGTSVPLSELTIYWTVDKQPVFNCDWAQVGCNVRGSGFKLSTGRTGADYY

TEITFTSGAGSLAAGASSGDIQVRINKNDWTNNEANDYSYDPTKTSFADNRVTLR

NGQLLWGVEP"

1. .60

/note="Region: proline-threonine linker"

/evidence=not\_experimental

61. .510

/note="Region: cellulose-binding domain type IIIB"

/evidence=not\_experimental

620. .3412

/gene="mana"

620. .3412

/gene="mana"

/codon\_start=1

/transl\_table=11

/product="multidomain beta-1,4-mannanase precursor"

/protein\_id="AAF22274.1"

/db\_xref="GI:6651327"

/translation="MNRRLIARLSGLAMVLIAMLAAYVKKPPEAAGHGMVFPATRT

YACYGVKGVHNGDLMINPAICLDAISGNYQFNWFCNGLISNAGGRHREIIPDGK

LCGTASFDGMNQARTDWTTRLPQCATITVRVNAHPHGTWLYVTRDGDWPTPL

KMSLEPTFPFQINSQDGAISQVQPNKGRHIIYIMIQWSDSPAFYN

CSDVYFGSGPIAYEFQDPREGGTMITPPSGTTPPTPTPTPTPTPTPTPTPTPTPT

TPSTPTPSASGTLRVYRVGDTSATDNQKPOLRIVNTGSQAVPLTELKVRVYTKN

STQAOYFCDMAQIGCSNRAQPVSLQSPVSGADSYIELSTGSGIPAGNGTGIQNR

IHFTWMNYNETDDSYNGAQNTWGPSTRIITLVNGVLVWGTEPGGSGSSPPTPTPT

GANAVRIVLSNGCRWTKIPASEVADIIISQARTIGYRAVLEVDHDTGIGDAAACMT

VPANWIELKNVLAGQENFVINGNEPYNNGYQNVVTDTRNAVQALRAGNNTIM

PLVIGPFGHSHSDPDNEQAIQVAKYNIIGLFWSGSGGSGVYELDMVTFNANSPL

TANGWFRNTAIGTSTPT

KGPHRVNRGTSVPLSELTIYWTVDKQPVFNCDWAQVGCNVRGSGFKLSTGRTG

ADYYIETFTSGAGSLAAGASSGDIQVRINKNDWTNNEANDYSYDPTKTSFADNR

RVTLVNGQLVWGVEP"

620. .3409

/gene="mana"

/note="encodes multidomain beta-1,4-mannanase precursor"

misc\_feature



```
/evidence=not_experimental
620. .718
/gene="mana"
/evidence=not_experimental
719. .3409
/gene="mana"
/product="multidomain beta-1,4-mannanase"
1349. .1480
/gene="mana"
/note="Region: proline-threonine linker"
/evidence=not_experimental
1481. .1927
/gene="mana"
/note="Region: cellulose-binding domain type IIb"
/evidence=not_experimental
1928. .2032
/gene="mana"
/note="Region: proline-threonine linker"
/evidence=not_experimental
2033. .2884
/gene="mana"
/note="Region: beta-1,4-mannanase catalytic domain"
/evidence=experimental
2885. .2959
/gene="mana"
/note="Region: proline-threonine linker"
/evidence=not_experimental
2960. .3409
/gene="mana"
/note="Region: cellulose-binding domain type IIb"
/evidence=not_experimental
3894. 4400
/note="orf3"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF22275.1"
/db_xref="GI:6651328"
/translation="MGASGVPSRTAONGYGLKARCQRCVGRNLLGHRLEFVDDDEF
RFFAGARKLRFRNRQNGEAGNFESGTVTFQDDKRRIRENIGQLFVRFOASI
GPKRLIEAQERQARQASDQPLKNAVILHFIDQHPYGVVERMPDFHFKQKS
RSILTRY"
ORIGIN
Alignment Scores:
Pred. No.: 1.91e-17 Length: 4567
Score: 225.50 Matches: 41
Percent Similarity: 67.44% Conservative: 17
Best Local Similarity: 47.67% Mismatches: 27
Query Match: 48.08% Indels: 1
DB: 1 Gaps: 1
US-09-917-376-5 (1-88) x AF163837 (1-4567)
Qy 3 GlyGlyVallysValGlnTyrIysAsnAsnApsSerAlaProGlyAspAsnGlnIlelys 22
Db 58 GCGACCTTTCGGATTCAGTATCGCGCGGCGGACACGAGCGGCGGCAACACAGCTGAAG 117
Qy 23 ProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrVal 42
Db 118 CCGCATTTTCGATTTGTAACCGCGGGACGTCGAGCGTCCGCTGTCGGAGCTGACGATT 177
Qy 43 ArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTirpala 62
Db 178 CGGTACTGGTACACGGTGCAC---GGGACAAAGCGGACGGTGTTCACCTGCGGCTGGCGC 234
Qy 63 AlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThr 82
Db 235 CAGTGGGTGTTCGAACGTCGCGGCGGACGCTTCGTGAAGCTTTCGACGGCGCGGACGGG 294
Qy 83 AlaAspThrTyrLeuGln 88
Db 295 CGGGACTACTATATCGAG 312
```

```
RESULT 8
AF200304
LOCUS
DEFINITION
Caldibacillus cellulovorans putative type IIb cellulose-binding
domain gene, partial cds; and beta-1,4-xylanase Xyna precursor
(xyna) gene, complete cds.
ACCESSION
AF200304
VERSION
AF200304.1 GI:7385018
SOURCE
Caldibacillus cellulovorans
ORGANISM
Bacteria; Firmicutes; Bacillales; Alicyclobacillaceae;
Caldibacillus.
REFERENCE
1 (bases 1 to 3237)
Sunna,A., Gibbs,M.D. and Bergquist,P.L.
A novel thermostable multidomain 1,4-beta-xylanase from
'Caldibacillus cellulovorans' and effect of its xylan-binding
domain on enzyme activity
Microbiology 146 (Pt 11), 2947-2955 (2000)
JOURNAL
MEDLINE
20519260
PUBMED
11065373
REFERENCE
2 (bases 1 to 3237)
Sunna,A., Gibbs,M.D. and Bergquist,P.L.
Direct Submission
TITLE
Submitted (29-OCT-1999) Biological Sciences, Macquarie University,
North Ryde, Sydney, New South Wales 2109, Australia
JOURNAL
FEATURES
Location/Qualifiers
1. .3237
/organism="Caldibacillus cellulovorans"
/mol_type="genomic DNA"
/db_xref="taxon:74586"
<1. .279
/note="orf1"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative type IIb cellulose-binding domain"
/protein_id="AAF61648.1"
/db_xref="GI:7385019"
/translation="QVGCNLRGSFVKLSTGRGTADYIEITFTSGAGSLAPGASGSD
IQVRINKNDWTNYNEANDYSYDPTKTSFADNRVTLRYNGQLIWGVEP"
384. .3149
/gene="xyna"
384. .3149
/gene="xyna"
/note="multidomain protein"
/codon_start=1
/transl_table=11
/product="beta-1,4-xylanase Xyna precursor"
/protein_id="AAF61649.1"
/db_xref="GI:7385020"
/gene="xyna"
384. .482
/gene="xyna"
483. .3146
/gene="xyna"
sig_peptide
mat_peptide
```



[illegible]



```
Qy 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
Db 2999 GTTCGTTATTATTTTCAAGAGGAC--AGCGCAGCGGACATGAGCTTCTGGTGTGACTAT 3055
Qy 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81
Db 3056 GCACAGCTTGGCAGTGCCTAATGTGAGCGGACGCTTGTGCTGTAAATCCGCGGAAGGCG 3115
Qy 82 ThrAlaAspThrTyrLeuGln 88
Db 3116 ACGGCAGATACGTATCTCGAG 3136

RESULT 10
BSPI33614
LOCUS BSPI33614 Bacillus sp. BP23 celB gene. 4161 bp DNA linear BCT 09-MAY-2001
DEFINITION AJ133614
ACCESSION AJ133614.1 GI:4490765
VERSION celB gene; cellulase; glycosyl hydrolase.
KEYWORDS Bacillus sp. BP-23
SOURCE Bacillus sp. BP-23
ORGANISM Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

REFERENCE
1
Pastor,F.I., Pujol,X., Blanco,A., Vidal,T., Torres,A.L. and Diaz,P.
Molecular cloning and characterization of a multidomain
endoglucanase from Paenibacillus sp BP-23: evaluation of its
performance in pulp refining
Appl. Microbiol. Biotechnol. 55 (1), 61-68 (2001)
MEDLINE 21129642
PUBMED 11234960
REFERENCE 2 (bases 1 to 4161)
Diaz,P.
Direct Submission
TITLE Submitted (10-MAR-1999) Diaz P., Microbiology, University of
Barcelona, Av.Diagonal 645, Barcelona-08028, SPAIN
JOURNAL Location/Qualifiers
1. 4161
/organism="Bacillus sp. BP-23"
/mol_type="genomic DNA"
/db_xref="taxon:89769"
398..3525
/gene="celB"
398..404
/gene="celB"
437..442
/gene="celB"
514..519
/gene="celB"
532..3525
/gene="celB"
/EC_number="3.2.1.4"
/standard_name="endoglucanase"
/codon_start=1
/transl_table=11
/product="cellulase"
/protein_id="CAB38941.1"
/db_xref="GI:4490766"
/db_xref="GOA:Q92411"
/db_xref="UniProt/TREMBL:Q92411"
/tranlation="MKGSWRRVAIVTLISAGLLAGSPSMNTGLRQADAAAGDHNYAEA
LOKALYFEARSGPLPANNRVERGDSGMQDGVLDLTCGWYDAGDHVKGFPNHA
ASATLAWSVVEYADGYEQAGOLEIKDNIWATDYFMKAHTKPNELMGQVAGNTH
AWNGPAMQMNRFSEFKIDASCPSGDLAAETAALAAASSIVFADSDPYVSKLQCHAK
ELYNFADTYRGKYDCTIDAAAFYNSWTGYEDLAWGAWLYLATNDNAYLSKLSAA
DRWSTSGSANWPTVTCGWDPSKHVGAOILLARTISLNMPEATKFIQSTERNLDYWT
VCTNGRVRKYTPGGLAWLDQWGLRLYANAAPISFYVSDVSDVPVKSRYNPATSOI
NYILGDNRQSVYVGYGQSPQPHHRTAHSSWNNEIDPANHRHLLYLGAWGQPNHA
SDQYDDIGDYSVEVATDYTAGFTGALAKNLLYQGNHPLANFPAPPEVKGDFYVE
AAVRSSGNSYTEIRALLNRRSGWPARMGDQLSFYFLDLSVYAGRTVSDVQVTVS
SEGATVSDPVVDAAKRIYATANPSNTKIYPGGEGNYRKEVQFRIITPQCGAWNPAND
PSYONLTGNSVKSNYIPVDYAGVKVSGEPQVTPVAPPAAGVQAVAGNSVALNW
SASAGASVYTKRAEVSQGPYTTVAAGVNGLYTNTGLTNGKTYYYVVVTVNSAGSP
```

```
ASVQVSGMPQAAATTPGAVTLTSGTAGNNQNMNLSWTAASGASSYTVQRAVAGTGTDDVA
TGLAVLNTDTTALNGTSYSYRAAVNASGQTLNSNVMILTPSGPVPSTGLTLEQVRSRG
SGNSNNATVPOFNKNTGTQAIIDLSTVKIRYFTKDGTEELSPFDCWDAQVGSANVOG
MFVAVNPAGKTADTVREISFTSGAGSLAAGAETGVIQTRFSKNNSAFDQSNDSYDA
SKTAPAAWKVTAIQGNTQVWGLEP"
532..633
/gene="celB"
/organism="celB"

ORIGIN
Alignment Scores: 7.03e-15 Length: 4161
Pred. No.: 204.50 Matches: 38
Score: 204.50 Conservative: 21
Percent Similarity: 67.82%
Best Local Similarity: 43.68% Mismatches: 27
Query Match: 43.60% Indels: 1
DB: 1 Gaps: 1

US-09-917-376-5 (1-88) x BSPI33614 (1-4161)
Qy 2 SerGlyGlyValIysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
Db 3067 ACAGGTACGCTTGAGTGCAGTATCGAAGCGGGGTTGAGGCAATCCAGCAATGCAGTT 3126
Qy 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrValThr 41
Db 3127 ACTCGCAATTCAAATCTCAAGAATATACCGGCACACAGCGGATGATCTGTAGTACGGTGA 3186
Qy 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
Db 3187 ATTCTGTTATTACTTTTACCAAGAC---GGCAGCGAGGAGCTCTCTTCTGGTGTGATTAT 3243
Qy 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81
Db 3244 GCGCAGGTGGGCTCAGCGAAGTACAGCAATGTTTGTAGCAGTGAATCCGCGAAGGT 3303
Qy 82 ThrAlaAspThrTyrLeuGln 88
Db 3304 ACACGGGATACGTATGTGGAG 3324

RESULT 11
E35142
LOCUS E35142 2029 bp DNA linear PAT 18-JUN-2001
DEFINITION Truncated cellulase composition.
ACCESSION E35142
VERSION E35142.1 GI:13018967
KEYWORDS JP 1999221086-A/44.
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1 (bases 1 to 2029)
AUTHORS Paiji,A., Petaer,L.B., Roy,M.D., Gurahamu,K.F., Moreland,D.G.,
Hyu,M. and Dalan,P.W.
TITLE Truncated cellulase composition
JOURNAL Patent: JP 1999221086-A 44 17-AUG-1999;
CLARIANT INTERNATIONAL LTD
COMMENT OS Artificial Sequence
PN JP 1999221086-A/44
PD 17-AUG-1999
PF 21-SEP-1998 JP 1998283606
PI 19-SEP-1997 US 08/932571
PI PAIJI ANDERSON,PETAER L BAGUKUISUTO,ROY M DANIEL, PI
GURAHAMU K PARINTON,
PI MORELAND DAVID GIBUSU,HYU MORGAN,DAIAN PURATONIORISU WILLIAM
PC C12N15/09,C11D3/386,C12N1/21,C12N9/42//C12N1/21,C12R1:19, PC
(C12N9/42,C12R1:19),C12N15/00
CC
FH Key Location/Qualifiers
FT source 1..2029 /organism='Artificial Sequence'.
FT Location/Qualifiers
1..2029
/organism="synthetic construct"
/mol_type="genomic DNA"

FEATURES
source
```



/db\_xref="taxon:32630"

## ORIGIN

Alignment Scores: 3.78e-14 Length: 2029  
Pred. No.: 195.50 Matches: 38  
Score: 195.50  
Percent Similarity: 60.23% Conservativity: 15  
Best Local Similarity: 43.18% Mismatches: 34  
Query Match: 41.68% Indels: 1  
DB: 6 Gaps: 1

US-09-917-376-5 (1-88) x E35142 (1-2029)

```
QY 1 ValSerGlyGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
DB 1 ATGGGAAGTGGTGTGAAGTACTGTACAAAGCAATGAGCAAGTGCAGCAGGTTCT 60
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrVal 40
DB 61 ATAGGCGCGTGTAAAGATAGTGAATGGAGGCGCAGCAGTGTGATCTTAGCAGGTT 120
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
DB 121 AGATAAGACTACTGGTACACAGTGGTGTGACAAAGCCACAGAGTGCAGTA---TGTGAC 177
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
DB 178 TGGGCACAGATAGGGGCAAGCAATGTGACATTTTGTGAAGCTTAGCAGCGGAGTG 237
QY 81 ProThrAlaAspThrTyrLeuGln 88
DB 238 AGTGAGCGGATTATTACCTGGAG 261
```

## RESULT 12

```
E35143 LOCUS 2029 bp DNA linear PAT 18-JUN-2001
DEFINITION Truncated cellulase composition.
ACCESSION E35143
VERSION E35143.1 GI:13018968
KEYWORDS JP 199221086-A/45.
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1 (bases 1 to 2029)
AUTHORS Paiji,A., Peters,L.B., Roy,M.D., Gurahamu,K.F., Moreland,D.G.,
Hyu,M. and Dajan,P.W.
TITLE Truncated cellulase composition
JOURNAL Patent: JP 199221086-A 45 17-AUG-1999;
COMMENT CLARIANT INTERNATIONAL LTD
OS Unidentified
PN JP 199221086-A/45
PD 17-AUG-1999
PF 21-SEP-1998 JP 199221086-A/45
PR 19-SEP-1997 US 08/932571
PI PAIJI ANDERSON,PETER L BAGKUISUTO,ROY M DANIEL, PI
GURAHAMU K FARINTON,
PI MORELAND DAVID GIBUSU,HYU MORGAN,DAIAN PURATONIORISU WILLIAM
PC C12N15/09,C11D3/386,C12N1/21,C12N9/42//C12N1/21,C12R1:19), PC
(C12N9/42,C12R1:19),C12N15/00
```

```
CC
FH Key Location/Qualifiers
FT source 1..2029
FT Location/Qualifiers
1..2029 /organism='Unidentified'.
```

## FEATURES

source  
1..2029 /organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"

## ORIGIN

Alignment Scores: 3.78e-14 Length: 2029  
Pred. No.: 195.50 Matches: 38  
Score:

```
Percent Similarity: 60.23% Conservativity: 15
Best Local Similarity: 43.18% Mismatches: 34
Query Match: 41.68% Indels: 1
DB: 6 Gaps: 1

US-09-917-376-5 (1-88) x E35143 (1-2029)
QY 1 ValSerGlyGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
DB 1 ATGGGAAGTGGTGTGAAGTACTGTACAAAGCAATGAGCAAGTGCAGCAGGTTCT 60
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrVal 40
DB 61 ATAGGCGCGTGTAAAGATAGTGAATGGAGGCGCAGCAGTGTGATCTTAGCAGGTT 120
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
DB 121 AAGATAAGACTACTGGTACACAGTGGTGTGACAAAGCCACAGAGTGCAGTA---TGTGAC 177
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
DB 178 TGGGCACAGATAGGGGCAAGCAATGTGACATTTTGTGAAGCTTAGCAGCGGAGTG 237
QY 81 ProThrAlaAspThrTyrLeuGln 88
DB 238 AGTGAGCGGATTATTACCTGGAG 261

RESULT 13
AF078038S5 LOCUS 6005 bp DNA linear BCT 11-FEB-2001
DEFINITION Caldicellulosiruptor sp. Tok7B.1 glycosyl hydrolase 6 gene, partial
cds; and CelE gene, complete cds.
ACCESSION AF078042
VERSION AF078042.1 GI:12743878
KEYWORDS 5 of 5
SOURCE Caldicellulosiruptor sp. Tok7B.1
ORGANISM Caldicellulosiruptor sp. Tok7B.1
Bacteria; Firmicutes; Clostridia; Clostridiales;
Syntrophomonadaceae; Caldicellulosiruptor.
REFERENCE 1 (bases 1 to 6005)
AUTHORS Gibbs,M.D., Reeves,R.A., Farrington,G.K., Anderson,P.,
Williams,D.P. and Bergquist,P.L.
TITLE Multidomain and multifunctional glycosyl hydrolases from the
extreme thermophile caldicellulosiruptor isolate Tok7B.1
JOURNAL Curr. Microbiol. 40 (5), 333-340 (2000)
MEDLINE 20171169
PUBMED 10706665
REFERENCE 2 (bases 1 to 6005)
AUTHORS Gibbs,M.D., Reeves,R.A., Farrington,G.K., Anderson,P.,
Williams,D.P. and Bergquist,P.L.
TITLE Direct Submission
JOURNAL Submitted (27-OCT-1999) Biological Sciences, Macquarie University,
Sydney, NSW 2109, Australia
FEATURES
source
1..6005 Location/Qualifiers
/organism="Caldicellulosiruptor sp. Tok7B.1"
/mol_type="genomic DNA"
/isolates="Tok7B.1"
/db_xref="taxon:80339"
<1..506
/notes="ORF6; multidomain"
/codon_start=3
/transl_table=11
/product="glycosyl hydrolase 6"
/protein_id="AAK06393.1"
/db_xref="GI:12743884"
/translations="GQDPTWNGTYTGNPNLHVKIDYXGDLGITASLANALLYSAAT
KEYVSDAEAKNLAKELLDMMWNLVDRDDKGLSAPKRGDKYRFEQVYIPAGWTGKM
PNGDVIKSGVKFIDIRSKYQDPDQKLVSAYNAGEAPEPRFHRFWAQCDIAIANTY
EILFNGQ"
<1..503
/misc_feature
/note="Region: family 48 glycosyl hydrolase domain"
```



CDS	617..5872	Db	2708	TGTTTAAAGATAGTAGTAATGGAGCGCAGCAGCTGTTGATCTTTAGCAGGGTTAAGATAAGA	2767
	/note="multidomain glycosyl hydrolase"	Qy	44	TyrTrpPheThrArgAspGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla	63
	/codon_start=1	Db	2768	TACTGTACACAGCTGATGGTGACAAAGCCACAGAGTGGCGTA---TGTGACTGGGCACAG	2824
	/evidence=experimental	Qy	64	MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla	83
	/transl_table=11	Db	2825	ATAGGCGAAGCAATGTGACATTCATTTTGTGAAGCTTTAGCAGCGGAGTGAGTGAGCGC	2884
	/product="CeleB"	Qy	84	AspThrTyrLeuGln	88
	/db_xref="GI:12743885"	Db	2885	GATTATTACCTGGAG	2899
	/protein_id="AAK06394.1"	RESULT 14			
	/translation="MQSMKAIKRVVSVITALLVLTSLCFPGIMPVKAVAGGTNYVGEA	E35100			
	LQKTIMEYFQMSKGLFVSKQLHEANDYFVKCHPSKIVYQVGDPTVDH	LOCUS	E35100	6416 bp	DNA
	NFWGPAEYQMQRPAKCDLSNPASSVVAETAFLAVSVIKERNQKAAASYLOHAK	DEFINITION	E35100	Truncated cellulase composition.	linear
	DLFEFADTTRADAGTATGTYTSGGFTDDLGAAVMLYIATNDSYLLTAAEELMSEY	ACCESSION	E35100		PAT 18-JUN-2001
	ANGNTYTQCDVDRVYGLIMLAKITGKELYGAVERNLDHWDRIYTPKGMAYLGT	VERSION	E35100.1	GI:13018925	
	WGSRYATTAFLACVADMGCDNSKTKYLNFAKSOIDVALGSGTSFVVGRTNY	KEYWORDS	JP 199221086-A/2.		
	POHPRHNSHWANSMKIPEYHRHILYALVGFGSGDDSYNDITDYQNEVACDYN	SOURCE	unidentified		
	AGIVGALAKMTQLYGGPIDDFKALETPTDEIFVESFGNSQGPNTYEVLSIYNRT	ORGANISM	unclassified.		
	GWPRVTDKLSFKFYDITELIOAGYSPDVVKVDTYIEGKISGPVYMDKRNIIYY	REFERENCE	1 (bases 1 to 6416)		
	LVPFSGTKIYPGGEVEHKQAKPISVPQGPWDPNDPSYKGLTSOLEKNKYIAAYD	AUTHORS	Paiji, A., Petaer, L.B., Roy, M.D., Gurahamu, K.F., Moreland, D.G.,		
	YKNETSASTGISRPWKIYNGSGSSVDLSRVKIRYMYVDGDKPQSAVCMDAOIGAS	TITLE	Hyu, M. and Daiian, P.W.		
	NVTNFVLSGSGGADYILEVFGSSGAGQOPKDGDIQVRFNKNDSYNNQADDW	JOURNAL	Truncated cellulase composition		
	SWLQSMYNGNAKTYLVDGLVWGQEPGATPAPTSTATPTPTPTPTPTPTPTPT	COMMENT	Patent: JP 199221086-A 2 17-AUG-1999;		
	TPTVTSATPTAPTASPVGSYTPESYSGALKVYANGNLSSPTNVLNPKIENVGT		CLARIANT INTERNATIONAL LTD		
	TAVDLRSVKRYWTIDGEATQSVSVASSINPAVIDVKFVKLGANAGADYYVIGFK		OS Unknown		
	SGAGVLAAGQSTKIRLSIOKSGSYNOSNDYSVRSANSYIENKVTGYIDDVLVWGR		PN JP 199221086-A/2		
	EPGRNAQIKWYANGNLGSMTNVLPKIKIENVGTTAVDLRSVKRYWTIDGEATQS		PD 17-AUG-1999		
	VSVTSINPAVIDVKFVKLGANAGADYYVIGFKSGAGVLAAGQSTKIRLSIOKSG		PF 21-SEP-1998		
	GSYNOSNDYSVRSATGYIENKVTGYIDGAIWGRSPRGTKPAGGTPPTPAPTST		PR 19-SEP-1997		
	PPT		PI PAIJI ANDERSON, PETAER L BAGUKUISUTO, ROY M DANIEL, PI		
	PARLLGNRLGYNNWNNMNSNAGSDWYHSSDDPYCYIMGITGNDKNVPAVAVSKPHEQ		GURAHAMU K FARINTON,		
	SIKONAYSAILTQWVYVAKDNGTVSESETAPSPRAEVRKFKDGALSLQDPNDNY		PI MORELAND DAVID GIBUSU, HYU MORGAN, DAIAN PURATONITOTISU WILLIAM		
	YMDPEFYNLINKYGRSSATGKIGYILNDPEDLWFTTHPRHHPQKVTCSELINKSVE		PC C12N15/09, C11D3/386, C12N1/21, C12N9/42/(C12N1/21, C12R1:19), PC		
	LAKVKTLDPAEIFGPASYGVGLTLQDAPDNQVKNRHLFSLWLEQMKKASDS		(C12N9/42, C12R1:19), C12N15/00		
	FGKRLDVLDDHWYEAQVGVRCIFGDNSTSRDAIARMQAPRTLWDPTKYTKQKG				
	QITAGENSWINWPEYLPPLNPKADIDKYPTGKLAITEFDYGGKDHISGGLALAD				
	VLGIFGKYGVMAARWGDGSGYAOAAVNYILNYDKGSRGYSTCVSAETDVNNMPVY				
	ASIGEDDSTVHILINRNYDRKLKAEIKMNRVYTTGGEIYGFPDSTSSQIRKMGVLS				
	NIQNTTITIEVNTLVTHIVLTSSK"				
	725..2053				
misc_feature	/note="Region: family 9 glycosyl hydrolase domain"				
misc_feature	2054..2545				
misc_feature	/note="type Iiic cellulose binding site"				
misc_feature	2627..3106				
misc_feature	/note="type Iib cellulose binding site"				
misc_feature	3224..4195				
misc_feature	/note="duplicated type IiIb cellulose binding site"				
misc_feature	4298..5869				
misc_feature	/note="Region: family 44 glycosyl hydrolase domain"				
misc_feature	2537..2626				
misc_feature	/note="PT-linker"				
misc_feature	3107..3223				
misc_feature	/note="PT-linker"				
misc_feature	4196..4295				
misc_feature	/note="PT-linker"				
ORIGIN					
Alignment Scores:					
Pred. No.:	1.96e-13	Length:	6005		
Score:	194.50	Matches:	38		
Percent Similarity:	61.18%	Conservative:	14		
Best Local Similarity:	44.71%	Mismatches:	32		
Query Match:	41.47%	Indels:	1		
DB:	1	Gaps:	1		
US-09-917-376-5 (1-88) x AF078038S5 (1-6005)					
Qy	4 GlyVallysValGlnTyrIlyAsnAsnAspSerAlaProGlyAspAsnGlnIleLysPro	23			
Db	2648 GGTGTGAAGGTACTGTACAAGAACAAATGAGACAAAGTGGCAGCACAGGTTCTATAAGCGCG	2707			
Qy	24 GlyLeuGlnLeuValAsnThrClySerSerValAspLeuSerThrValThrValArg	43			
Db	2725 TGGTTTAAAGATAGTAGTAATGGAGCGCAGCAGCTGTTGATCTTTAGCAGGGTTAAGATAAGA	2784			
Qy	44 TyrTrpPheThrArgAspGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla	63			
Db	2785 TACTGTACACAGCTGATGGTGACAAAGCCACAGAGTGGCGTA---TGTGACTGGGCACAG	2841			







GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model  
Run on: October 5, 2005, 06:17:10 ; Search time 135.381 Seconds  
(without alignments)  
3847.935 Million cell updates/sec

Title: US-09-917-376-5  
Perfect score: 469  
Sequence: 1 VSGGVKQYKNDSPAGDQ.....IRASFGSVNPATPTADTYLQ 88

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues  
Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame\_p2n.model -DEV=xlh  
-O=/cpn2.1/USPTO.spool/US09917376/runat\_04102005\_164327\_26889/app\_query.fasta\_1.2524  
-DB=N\_Geneseq\_16Dec04 -QWTF=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09917376 @CGN 1 1 1867 @runat\_04102005\_164327\_26889 -NCPU=6 -ICPU=3  
-NO WMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOF=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_16Dec04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004as:\*  
13: geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	469	100.0	2289	8 ABZ77633	Abz77633 Nucleotid
2	469	100.0	2289	10 ADJ38292	Adj38292 A. cellu
3	466	99.4	3687	8 ABZ77634	Abz77634 Nucleotid
4	466	99.4	3687	12 ADH36637	Adh36637 DNA encod
5	462	98.5	2869	10 ADD22922	Add22922 Acidother

6	462	98.5	2869	10 ABZ77632	Abz77632 Nucleotid
7	462	98.5	2869	12 ADOS2314	Ados2314 Acidother
8	436	93.0	3365	8 ABZ76162	Abz76162 A. cellu
9	409	87.2	3364	10 ADF75896	Adf75896 Acidother
10	231.5	49.4	2600	2 AAQ15178	Aaq15178 Portion o
11	195.5	41.7	2029	2 AAX55660	Aax55660 DNA seque
12	195.5	41.7	2029	6 AAD26568	Aad26568 Active ce
13	194.5	41.5	6415	2 AAX55662	Aax55662 DNA seque
14	194.5	41.5	6416	6 AAD26526	Aad26526 Active ce
15	192.5	41.0	11706	2 AAX55661	Aax55661 DNA seque
16	192.5	41.0	11707	6 AAD26525	Aad26525 Active ce
17	190	40.5	1527	13 ADT48400	Adt48400 Bacterial
18	190	40.5	2177	11 ADOS5906	Ados5906 Bacillus
19	188	40.1	1434	6 AAL41028	Aal41028 CMCase ge
20	188	40.1	1488	6 AAL41025	Aal41025 CMCase ge
21	188	40.1	2510	6 ABK53202	Abk53202 Bacillus
22	186	39.7	2175	2 AAQ49820	Aaq49820 NK-1 cell
23	182.5	38.9	2977	2 AAQ13001	Aaq13001 Endol gen
24	165	35.2	501	10 ADG14263	Adg14263 cellulose
25	165	35.2	717	10 ADG14257	Adg14257 cellulose
26	165	35.2	1438	3 AAZ45336	Aaz45336 DNA encod
27	165	35.2	1482	2 AAX90978	Aax90978 DNA encod
28	165	35.2	1482	2 AAZ31562	Aaz31562 Pectate l
29	165	35.2	5562	2 AAT86625	Aat86625 C. thermo
30	163	34.8	1500	12 ADJ35111	Adj35111 DNA encod
31	157	33.5	1314	6 ABK73393	Abk73393 Bacillus
32	157	33.5	2602	12 ADG32259	Adg32259 DNA encod
33	155.5	33.2	4059	5 AAH75059	Aah75059 Nucleotid
34	150.5	32.1	4056	5 AAH75060	Aah75060 Nucleotid
35	130.5	27.8	4452	8 ACA27429	Acta27429 Prokaryot
36	126.5	27.0	1303	10 ADC27473	Adc27473 Fusion pr
37	126.5	27.0	1747	10 ADC27474	Adc27474 Fusion pr
38	126.5	27.0	3489	10 ADC27475	Adc27475 Fusion pr
39	120.5	25.7	1775	2 AAQ13003	Aaq13003 Endo3 gen
40	114	24.3	486	2 AAQ72917	Aaq72917 Cellulose
41	114	24.3	486	2 AAQ74072	Aaq74072 C. cellul
42	114	24.3	499	2 AAX24930	Aax24930 Clostridi
43	114	24.3	507	5 AAD11042	Aad11042 Clostridi
44	114	24.3	768	4 AAF86248	Aaf86248 DNA seque
45	114	24.3	768	6 ABK52403	Abk52403 E. coli c

ALIGNMENTS

RESULT 1  
ABZ77633  
ID ABZ77633 standard; DNA; 2289 BP.  
XX  
AC ABZ77633;  
XX  
DT 03-JUN-2003 (first entry)  
XX  
DE Nucleotide sequence of the ManA polypeptide.  
XX  
KW ManA; mannanase A; glycoside hydrolase; enzyme; hemicellulose; sugar;  
KW food; feed; paper pulp; biofuel; manase; gene; ss.  
XX  
OS Acidothermus cellulolyticus.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..2289  
FT /\*tag= a  
FT /product= "ManA"  
XX  
PN WO2003012110-A1.  
XX  
PD 13-FEB-2003.  
XX  
PF 28-JUL-2001; 2001WO-US023819.  
XX  
PR 28-JUL-2001; 2001WO-US023819.  
XX  
XX (WIDE ) MIDWEST RES INST.  
PA



```
XX Ding S, Adney WS, Vinzant TB, Himmel ME;
PI WPI; 2003-248182/24.
DR P-PSDB; ABP73022.
XX Novel thermal tolerant mannanase A polypeptide derived from Acidothermus
PT cellulolyticus, useful for reducing hemicellulose in a starting material,
PT for processing of food, and as bulking agents in food stuffs.
XX Example 1; Page 23; 46pp; English.
XX The present sequence encodes ManaA, a thermostable mannanase A polypeptide
CC derived from Acidothermus cellulolyticus. ManaA is a member of the
CC glycoside hydrolase family of enzymes. ManaA is useful for reducing
CC hemicellulose in a starting material to simpler carbohydrate units, and
CC ultimately to sugars which are useful in the food, feed, paper pulp, and
CC biofuels industries. It is useful for the processing of food and in food
CC stuffs as bulking agents, and for the degradation of mannanase. ManaA is
CC also useful to raise polyclonal and monoclonal antibodies that are useful
CC in purifying ManaA, or detecting ManaA polypeptide expression, and as well
CC as reagent tools for characterizing the molecular actions of ManaA
CC polypeptides
XX SQ Sequence 2289 BP; 463 A; 700 C; 688 G; 438 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 5.36e-48 Length: 2289
Score: 469.00 Matches: 88
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 8 Gaps: 0

US-09-917-376-5 (1-88) x ABZ77633 (1-2289)

QY 1 ValSerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
DB 1363 GTGTGGGTGGGTGAAGTGCAGTACAGAACATGATTGGCGCCGGGTGATACCAG 1422
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
DB 1423 ATCAAAACGGGTCTCCAGTTGGTGAATACGGGGTCTGTCGGTGGATTGTTCGACGGTG 1482
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
DB 1483 ACGTGGCGGTACTGGTTCCACCGGATGGTGGGTCTGTCGACACTGGTGTACAACTGTGAC 1542
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
DB 1543 TGGCGCGCGATGGGGTGTGGGAAATATCCGGCTCTGTTGGCTCGGTGAACCCGGCGAG 1602
QY 81 ProThrAlaAspThrTyrLeuGln 88
DB 1603 CCGACGGCGGACACCTACCTGCAG 1626

RESULT 2
ADJ38292
ID ADJ38292 standard; DNA; 2289 BP.
XX AC ADJ38292;
XX 06-MAY-2004 (first entry)
XX A. cellulolyticus mannanase ManaA DNA.
XX mannanase A; catalytic domain GH5; carbohydrate binding domain III;
KW carbohydrate binding domain II; food processing; foodstuff;
KW bulking agent; ManaA; ds; gene.
XX Acidothermus cellulolyticus.
XX Key Location/Qualifiers
FH
```



```
XX 03-JUN-2003 (first entry)
XX Nucleotide sequence of the GuxA polypeptide.
XX GuxA; cellulase; glycoside hydrolase; enzyme; cellulose; sugar; biofuel;
XX detergent; pulp processing; paper processing; feed processing; textile;
XX gene; 88.
XX Acidothermus cellulolyticus.
XX Key Location/Qualifiers
XX CDS 1..3687
XX FT /*tag= a
XX FT /product= "GuxA"
XX WO2003012109-A1.
XX 13-FEB-2003.
XX 28-JUL-2001; 2001WO-US023817.
XX 28-JUL-2001; 2001WO-US023817.
XX (WIDE ) MIDWEST RES INST.
XX PA
XX Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;
XX WPI; 2003-239526/23.
XX P-PSDB; ABP73029.
XX Novel thermal tolerant GuxA polypeptide derived from Acidothermus
XX cellulolyticus, useful for reducing cellulose in a starting material, and
XX for the conversion of biomass to biofuels and biofuel additives.
XX Example 1; Page 23-24; 47pp; English.
XX The present sequence encodes a GuxA polypeptide. GuxA is thermostable
XX cellulase, and is a member of the glycoside hydrolase family of enzymes.
XX GuxA is useful for reducing cellulose in a starting material such as
XX agricultural biomass to sugars. This is useful in biofuel production.
XX GuxA is also useful in the conversion of biomass to biofuels and biofuel
XX additives, in detergents, pulp and paper processing, food and feed
XX processing, and in textile processes. GuxA is also useful for raising
XX polyclonal and monoclonal antibodies that are useful in purifying GuxA,
XX or detecting GuxA polypeptide expression, as well as reagent tools for
XX characterizing the molecular actions of GuxA polypeptides
XX
XX Sequence 3687 BP; 725 A; 1173 C; 1171 G; 618 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 2,33e-47 Length: 3687
Score: 466.00 Matches: 87
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 98.86% Mismatches: 0
Query Match: 99.36% Indels: 0
Gaps: 8
DB:
US-09-917-376-5 (1-88) x ABZ77634 (1-3687)
QY 1 ValSerGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAenGln 20
Db 1750 GTGTGGGTGGGTGAAGGTGCAGTACAGAACATGATTCGGCGCGGTGATACCCAG 1809
QY 21 ILslyProGlyLeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrVal 40
Db 1810 ATCAAAACCGGGTCTCAGTTGGTGAATACCGGTCTGTCGGTGGATTGTGACGGTG 1869
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1870 ACGTGGCGGTACTGGTTTACCCCGGATGGTGGGTCTGTCACACTGGTGTACACTGTGAC 1929
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
```

```
Db 1930 TGGGCGCGATGGGTGTGGGAATATCCGCCCTGTTTCGGTCTCGTGAACCCGCGACG 1989
QY 81 ProThrAlaAspThrTyrLeuGln 88
Db 1990 CCGACGGCGGACACCTACTGCGAG 2013
RESULT 4
ADH36637
ID ADH36637 standard; DNA; 3687 BP.
XX
XX AC ADH36637;
XX DT 11-MAR-2004 (first entry)
XX DE DNA encoding Acidothermus cellulolyticus glycoside hydrolase, GuxA.
XX KW Thermal tolerant cellulase; glycoside hydrolase; GuxA;
XX KW carbohydrate degradation; cellulose; agricultural biomass;
XX KW municipal solid waste; thermostable; gene; ds.
XX OS Acidothermus cellulolyticus.
XX Key Location/Qualifiers
XX CDS 1..3687
XX FT /*tag= a
XX FT /product= "GuxA"
XX PN US2003104522-A1.
XX PD 05-JUN-2003.
XX PF 28-JUL-2001; 2001US-00917383.
XX PR 28-JUL-2001; 2001US-00917383.
XX (DING/) DING S.
XX (ADNE/) ADNEY W S.
XX (VINZ/) VINZANT T B.
XX (HIMM/) HIMMEL M E.
XX (DECK/) DECKER S R.
XX Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;
XX WPI; 2004-106451/11.
XX P-PSDB; ADH36636.
XX Novel isolated thermostable GuxA polypeptide useful for detecting
XX polynucleotide encoding GuxA, assessing carbohydrate degradation activity
XX of GuxA, reducing cellulose in starting material e.g., agricultural
XX biomass.
XX Claim 21; SEQ ID NO 2; 20pp; English.
XX The present invention relates to the isolation of a thermal tolerant
XX cellulase from Acidothermus cellulolyticus. The cellulase is a member of
XX the glycoside hydrolase family and is designated GuxA. Also disclosed are
XX methods of producing and using GuxA. The GuxA polypeptide is useful for
XX the detection of a polynucleotide encoding GuxA. The polypeptide sequence
XX is also useful for assessing the carbohydrate, e.g. cellulose,
XX degradation activity of GuxA. GuxA is useful for reducing cellulose in a
XX starting material such as agricultural biomass or municipal solid waste.
XX The polypeptide molecule of GuxA is thermostable. The present sequence
XX encodes A. cellulolyticus glycoside hydrolase, GuxA.
XX Sequence 3687 BP; 725 A; 1173 C; 1171 G; 618 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 2,33e-47 Length: 3687
Score: 466.00 Matches: 87
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 98.86% Mismatches: 0
Query Match: 99.36% Indels: 0
```



```
DB: 12 Gaps: 0
US-09-917-376-5 (1-88) x ADH36637 (1-3687)
QY 1 ValSerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 1750 GTGTCGGGTGGGCTCAAGGTGCAGTACAAACAATGATTCCGGCGCCGGGTGATAACACAG 1809
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1810 ATCAAAACGGGTCTCCAGTGTGTGATACCCGGTCTGTCGTGGATTTGTGACGGTG 1869
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1870 ACGGTGGGTACTGTTTCAACCGGGATGTTGGTGTGTCGACACTGTTGTACAACTGTGAC 1929
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 1930 TGGCGCGCGATGGGGTGTGGGAATATCCGCGCTCGTTCGGTCTCGTGAACCCGCGACG 1989
QY 81 ProThrAlaAspThrTyrLeuGln 88
Db 1990 CGACGCGCGACCTACCTCGAG 2013
RESULT 5
ADD22922
ID ADD22922 standard; DNA; 2869 BP.
XX
AC ADD22922;
XX
DT 15-JAN-2004 (first entry)
XX
DE Acidothermus cellulolyticus avicelase AvIII DNA.
XX
KW AvIII; cellulose reduction; agricultural biomass; municipal solid waste;
KW glycoside hydrolase; avicelase; ds; gene.
XX
OS Acidothermus cellulolyticus.
FH Key Location/Qualifiers
FT CDS 1..2869
FT /tag= a
FT /product= "AvIII"
XX
PN US2003108988-A1.
XX
PD 12-JUN-2003.
XX
PF 18-OCT-2002; 2002US-00155400.
XX
PR 28-JUL-2001; 2001US-00917376.
XX
PA (DING/) DING S.
PA (ADNE/) ADNEY W S.
PA (VINZ/) VINZANT T B.
PA (HIMM/) HIMMEL M E.
XX
PI Ding S, Adney WS, Vinzant TB, Himmel ME;
XX
WPI; 2003-810853/76.
DR P-PSDB; ADD22921.
XX
PT New isolated thermal tolerant avicelase polynucleotide useful for
PT detection of a polynucleotide encoding AvIII and for reducing cellulose
PT in a starting material, e.g. municipal solid waste.
XX
PS Claim 17; SEQ ID NO 2; 29pp; English.
XX
CC The invention relates to an isolated polynucleotide molecule encoding a
CC thermostable AvIII polypeptide. The polynucleotide is useful for
CC detection of a polynucleotide encoding AvIII. The polynucleotide is
CC useful for reducing cellulose in a starting material which involves
CC administering to the starting material, e.g. agricultural biomass or
CC
```

```
CC municipal solid waste, a polypeptide molecule of the polynucleotide. The
CC method further comprises administering a second polypeptide molecule
CC chosen from the glycoside hydrolase family of proteins. The present
CC sequence represents DNA encoding the Acidothermus cellulolyticus
CC avicelase AvIII.
XX
SQ Sequence 2869 BP; 546 A; 857 C; 900 G; 565 T; 0 U; 1 Other;
Alignment Scores: 5.3e-47 Length: 2869
Score: 462.00 Matches: 86
Percent Similarity: 100.00% Conservatives: 2
Best Local Similarity: 97.73% Mismatches: 0
Query Match: 98.51% Indels: 0
DB: 10 Gaps: 0
US-09-917-376-5 (1-88) x ADD22922 (1-2869)
QY 1 ValSerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 2605 GTGTCGGGTGGGCTCAAGGTGCAGTATAAGAATAATGATTCCGGCGCCGGGTGATAATCAG 2664
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 2665 ATCAAGCCGGTTCAGGTGTGTAATACCGGGTCTGTCGTGGATTTGTGACGGTG 2724
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 2725 ACGGTGGGTACTGTTTCAACCGGGATGTTGGTGTGTCGACACTGTTGTACAACTGTGAC 2784
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 2785 TGGCGCGCGATCGGGTGTGGGAATATCCGCGCTCGTTCGGTCTCGTGAACCCGCGACG 2844
QY 81 ProThrAlaAspThrTyrLeuGln 88
Db 2845 CCGACGCGCGACCTACCTCGAG 2868
RESULT 6
ABZ77632
ID ABZ77632 standard; DNA; 2869 BP.
XX
AC ABZ77632;
XX
DT 03-JUN-2003 (first entry)
XX
DE Nucleotide sequence of the avicelase AvIII.
XX
KW Avicelase; AvIII; glycoside hydrolase; enzyme; cellulase; biofuel;
KW detergent; pulp processing; paper processing; feed processing; textile;
KW cellulose; gene; ss.
XX
OS Acidothermus cellulolyticus.
FH Key Location/Qualifiers
FT CDS 1..2869
FT /tag= a
FT /partial
FT /product= "AvIII"
FT /transl_except= (pos:2869,aa:Xaa)
FT /note= "Xaa is an unspecified residue"
XX
PN WO2003012090-A2.
XX
PD 13-FEB-2003.
XX
PF 28-JUL-2001; 2001WO-US023818.
XX
PR 28-JUL-2001; 2001WO-US023818.
XX
PA (MIDE ) MIDWEST RES INST.
XX
PI Ding S, Adney WS, Vinzant TB, Himmel ME;
```







Db 2785 TGGCGCGGATCGGGTGGGAATATCGCGCTCGTTCGGTGAACCCGCGGACG 2844  
 |||...|||

QY 81 ProThrAlaAspThrTyrLeuGln 88  
 |||...|||

Db 2845 CCGACGGCGGACACCTACCTGCAG 2868  
 |||...|||

RESULT 8  
 ABZ76162

ID ABZ76162 standard; DNA; 3365 BP.

XX AC ABZ76162;

XX 29-MAY-2003 (first entry)

XX DE A. cellulolyticus Gux1 protein encoding DNA.

XX Gux1; thermal tolerant; exoglucanase; glycoside hydrolase; cellulose;

XX biofuel; detergent; pulp; paper processing; feed processing; textile;

XX cellulase; gene; ds.

XX OS Acidothermus cellulolyticus.

XX Key Location/Qualifiers

XX CDS 1..3365

XX /tag= a

XX /product= "Gux1"

XX /trans\_except= (pos: 682..683, aa: Pro)

XX /note= "this codon has an apparent one nucleotide

XX basepair deletion which alters the reading frame"

XX W02003012095-A1.

XX 13-FEB-2003.

XX 28-JUL-2001; 2001WO-US023820.

XX 28-JUL-2001; 2001WO-US023820.

XX (MIDE ) MIDWEST RES INST.

XX Adney WS, Ding S, Vinzant TB, Himmel ME, Decker SR;

XX Lantz Mccarter S;

XX WPI; 2003-300494/29.

XX P-PSDB; ABP71656.

XX New thermal tolerant Gux1 peptide having specified amino acid sequence,

XX useful in the degradation of cellulose to biofuels.

XX Disclosure; Page 22-23; 44pp; English.

XX The invention relates to a thermal tolerant Gux1 peptide from A.

XX cellulolyticus. The Gux1 exoglucanase is a member of the glycoside

XX hydrolase family and comprises a catalytic domain GH48, carbohydrate

CC binding domain type III, and a carbohydrate binding domain type II. The

CC polypeptide is useful in the degradation of cellulose into biofuel, or

CC for conversion of biomass to biofuel additives. It is used in detergents,

CC pulp and paper processing, food and feed processing, and in textile

CC processing. It can also be used alone or in combination with other

CC cellulase or glycoside hydrolases. The novel polypeptide generates

CC alternative cellulase enzymes capable of assisting in the commercial-

CC scale processing of cellulose to sugar for use in biofuel production. The

CC present sequence represents a A. cellulolyticus Gux1 cellulase encoding

CC DNA

XX SQ Sequence 3365 BP; 669 A; 1117 C; 1041 G; 538 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.11e-43 Length: 3365

Score: 436.00 Matches: 81

Percent Similarity: 97.62% Conservative: 1

Best Local Similarity: 96.43% Mismatches: 2

Query Match: 92.96% Indels: 0  
 DB: 8 Gaps: 0

US-09-917-376-5 (1-88) x ABZ76162 (1-3365)

QY 5 VallysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGlnIleLysProGly 24  
 ::|||

Db 112 CTCAAAGCGCAGTATAAGAACCAATGATTCGGCGCCGAGTGACCAACCAACCGGGT 171  
 |||

QY 25 LeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrValThrValArgTyr 44  
 |||

Db 172 CTCCAGTTGGTGAATACCGGGTCGTCGTCGGTGGATTTGTCGACCGGTGACGGTCCGGTAC 231  
 |||

QY 45 TrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaLaMet 64  
 |||

Db 232 TGGTTCACCCGGGATGGTGGTTCGTCGACACTGGTGTACAACTGTGACTGGGCGCGATG 291  
 |||

QY 65 GlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAlaAsp 84  
 |||

Db 292 GGGTGTGGGAATATCCGGCCTCGTTCCGCTCGGTGAACCCGCGGACGCCGCGCGGAC 351  
 |||

QY 85 ThrTyrLeuGln 88  
 |||

Db 352 ACCTACCTGCAG 363

RESULT 9

ADF75896

ID ADF75896 standard; DNA; 3364 BP.

XX AC ADF75896;

XX 26-FEB-2004 (first entry)

XX DE Acidothermus cellulolyticus Gux1 DNA.

XX Gux1; gene; ds; thermal tolerant; cellulase; glycoside hydrolase;

XX exoglucanase; cellulose; sugar; biofuel production.

XX OS Acidothermus cellulolyticus.

XX Key Location/Qualifiers

XX CDS 1..3364

XX /tag= a

XX /product= "Gux1 protein"

XX /note= "This polynucleotide sequence contains translation

XX exceptions"

XX US2003096342-A1.

XX 22-MAY-2003.

XX 28-JUL-2001; 2001US-00917384.

XX 28-JUL-2001; 2001US-00917384.

XX (ADNE/) ADNEY W S.

XX (DING/) DING S.

XX (VINZ/) VINZANT T B.

XX (HIMM/) HIMMEL M E.

XX (DECK/) DECKER S R.

XX (MCCA/) MCCARTER S L.

XX Adney WS, Ding S, Vinzant TB, Himmel ME, Decker SR, Mccarter SL;

XX WPI; 2003-863404/80.

XX P-PSDB; ADF75895.

XX Novel thermal tolerant cellulase of glycoside hydrolase family,

XX comprising catalytic domain, first and second carbohydrate binding

XX domain, isolated from Acidothermus cellulolyticus, useful in cellulose

XX degradation.

XX Claim 21; SEQ ID NO 2; 20pp; English.



XX This invention relates to a novel thermal tolerant cellulase, a member of  
CC the glycoside hydrolase family. Specifically, it refers to an  
CC exoglucanase identified as Gux1, and isolated from *Acidothermus*  
CC *cellulolyticus*, which comprises the catalytic domain GH48, and the  
CC carbohydrate binding domains type II and type III. The present invention  
CC describes Gux1 compositions that are useful for reducing cellulose in a  
CC starting material (or biomass). As such, these enzymes are capable of  
CC assisting in commercial scale processing of cellulose to sugar, for use  
CC in biofuel production. Furthermore, heterologous cellulase proteins (or  
CC fusion proteins) can be generated with specific desired properties  
CC including thermal stability and partial or complete resistance to extreme  
CC pH inactivation, proteolytic, solvent, detergent or chaotropic agent  
CC inactivation. This polynucleotide sequence is the *Acidothermus*  
CC *cellulolyticus* Gux1 DNA of the invention.

XX SQ Sequence 3364 BP; 669 A; 1110 C; 1048 G; 536 T; 1 U; 0 Other;

Alignment Scores:  
Pred. No.: 2.52e-40 Length: 3364  
Score: 409.00 Matches: 80  
Percent Similarity: 96.43% Conservative: 1  
Best Local Similarity: 95.24% Mismatches: 3  
Query Match: 87.21% Indels: 1  
DB: 10 Gaps: 0

US-09-917-376-5 (1-88) x ADF75896 (1-3364)

QY 5 VallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysProGly 24  
DB 112 CTCGAAGCGCAGTATAGAACATGATTCGGCGCGAGTGACCAACAGATCAAAACCGGT 171  
QY 25 LeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrValThrValArgTyr 44  
DB 172 CTCACGTTGGTGAATACCGCGTCTGTCGTGGTGAUTG-TCGACGCTGACGCGTGGGTAC 230  
QY 45 TrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAlaMet 64  
DB 231 TGGTTTCAACCGCGGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 290  
QY 65 GlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAlaAsp 84  
DB 291 GGGTGGGGAATATCCGCGCTCGTTCCGCTCGGTGAACCCGCGCGACGCGCGGAC 350  
QY 85 ThrTyrLeuGln 88  
DB 351 ACCTACCTGCAG 362

RESULT 10  
AAQ15178  
ID AAQ15178 standard; DNA; 2600 BP.

XX AAQ15178;

XX 27-AUG-2003 (revised)  
XX 25-MAR-2003 (revised)  
XX 17-MAR-1992 (first entry)

XX Portion of pAEC-1 contg. cellulase AE-1 gene.

XX Detergents; pharmaceuticals; deinking; carboxymethylcellulose; ss.

XX *Aeromonas* sp.

XX Key Location/Qualifiers  
XX CDS 192..2540  
XX /\*tag= a  
XX /product= "cellulase AE-1"

XX JP03251174-A.

XX 08-NOV-1991.

XX

PF 28-FEB-1990; 90JP-00045465.  
XX  
PR 28-FEB-1990; 90JP-00045465.  
XX  
PA (OJIP ) OJI PAPER CO.  
XX  
XX WPI; 1991-373412/51.  
DR P-PSDB; AAR15625.  
XX  
XX Cellulase AE-1 for e.g. mfr. of pharmaceuticals and foodstuffs - of opt.  
PT pH when carboxymethylcellulose is used as substrate.  
XX  
XX Claim 3; Fig 2; 8pp; Japanese.  
XX  
XX The plasmid, pAEC 1, was prepd. ligating chromosomal DNA contg. the gene  
CC (obtd. from *Aeromonas*) into pUC18. The gene can be used to produce  
CC recombinant enzymes which is used for the effective utilis- ation of  
CC biomass resources and the mfr. of pharmaceuticals and foodstuffs, and  
CC also for the detergent and deinking of waste paper. (Updated on 25-MAR-  
CC 2003 to correct PA field.) (Updated on 27-AUG-2003 to correct OS field.)  
XX  
XX SQ Sequence 2600 BP; 687 A; 649 C; 700 G; 564 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 2.11e-18 Length: 2600  
Score: 231.50 Matches: 42  
Percent Similarity: 66.67% Conservative: 16  
Best Local Similarity: 48.28% Mismatches: 28  
Query Match: 49.36% Indels: 1  
DB: 2 Gaps: 1

US-09-917-376-5 (1-88) x AAQ15178 (1-2600)

QY 2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21  
DB 2082 AGCGGGGATTTGGCGGTTTCAGTACAAACCGGAGATACGAATCGACCGCAACACGTTT 2141  
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrValThr 41  
DB 2142 AAGCCTCATTTCAATATCGTGAACAGGGGGTTCGGCGAGTGCCTTTAAGCGAGCTGAGT 2201  
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61  
DB 2202 CTGCGATATTATTTCACAGCTGAC---GGCAATGACCAACTGCAATATGAGCTGCTGG 2258  
QY 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81  
DB 2259 GCCATGGTAGGCTGCTCCAACTGAAACGGGGCTTTCGTGAAAATGATCCGGGCAAAAGCG 2318  
QY 82 ThrAlaAspThrTyrLeuGln 88  
DB 2319 AACGCGGATACCTACTTGGAG 2339

RESULT 11  
AAQ55660  
ID AAQ55660 standard; DNA; 2029 BP.

XX AAQ55660;

XX 30-JUL-1999 (first entry)

XX DNA sequence encoding truncated cellulase Cel E3/B5.

XX Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;  
XX Cel 1/2/3; Cel E3/B5; laundry detergent; stain-removing;  
XX cotton-containing fabric; stonewashing; ss.

XX Unidentified.

XX EP921188-A2.

XX 09-JUN-1999.

XX



```

PF 15-SEP-1998; 98EP-00810919.
XX
PR 19-SEP-1997; 97US-00932571.
XX
XX (CLRN ) CLARIANT FINANCE BVI LTD.
XX
PI Anderson P, Bergquist PL, Daniels RM, Farrington GK, Gibbs MD;
PI Morgan H, Williams DP;
XX
XX WPI; 1999-315403/27.
DR
DR P-PSDB; AAY13494.
XX
XX New truncated cellulase proteins, useful in detergents and for producing
PT 'stonewashed' denim.
PT
XX
PS Disclosure; Page 41-42; 65pp; English.
XX
CC The invention relates to a recombinant cellulase active protein free of
CC proteinases of native thermophilic and alkaliphilic origin, comprising
CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,
CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-
CC length sequences, or functional equivalents. Cel B5 extends from amino
CC acid A1011 to P1424 or K1425 or N1426, and Cel B4/5 extends from amino
CC acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends from
CC amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel E1/2/3
CC extends from Y39 to G812, Cel E6 extends from amino acid V1233 to K1751
CC and the stability region extends from amino acid E482 to G635 in the
CC sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new
CC enzymes are useful in laundry detergent compositions to prevent or remove
CC staining, backstaining or graying, for use on cellulosic materials
CC including cotton-containing fabrics. They are especially useful for
CC preventing redeposition of colorant during stonewashing, and for
CC processing of textiles where cellulose breakdown is required. The new
CC truncated enzymes show reduced redeposition of dye compared to using non-
CC truncated cellulase compositions
XX
SQ Sequence 2029 BP; 650 A; 329 C; 520 G; 530 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 4.59e-14 Length: 2029
Score: 195.50 Matches: 38
Percent Similarity: 60.23% Conservative: 15
Best Local Similarity: 43.18% Mismatches: 34
Query Match: 41.68% Indels: 1
DB: 2 Gaps: 1

US-09-917-376-5 (1-88) x AAX55660 (1-2029)
QY 1 ValSerGlyGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
DB 1 ATGGGAGTGGTGTGAAGGTACTGTACAAGAACAAATGACACAAAGTGGCAGCAGGTTCT 60
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrVal 40
DB 61 ATAAGCCGCGTGTAAAGATAGTGAATGGAGCGCAGCAGTGTGTGATCTTAGCAGGTT 120
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
DB 121 AAGATAAGATAGTGGTACACAGTGGTGTGACAAAGCCACAGAGTGGCGGTA---TGTGAC 177
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
DB 178 TGGGCACAGATAGGGGCAAGCAATGTGACATTCATTTTGTGAAGCTTAGCAGCGGAGTG 237
QY 81 ProThrAlaAspThrTyrLeuGln 88
DB 238 AGTGGAGCGGATATTATACCTGGAG 261

RESULT 12
AAD26568
ID AAD26568 standard; DNA; 2029 BP.
XX
AC AAD26568;

```

```

XX 26-MAR-2002 (first entry)
DT
XX Active cellulase hybrid protein, E3/B5 DNA.
DE
XX Active cellulase protein; alkaliphilic; textile processing; proteinase;
KW detergent additive; stonewashed appearance; cotton-containing denim;
KW CelB5; thermophilic; commercial detergent; E3/B5 hybrid protein; ds.
XX
OS Unidentified.
XX
FH Key Location/Qualifiers
FT CDS 1..1851
FT FT /*tag= a
FT FT /product= "E3/B5 hybrid protein"
XX
XX US6294366-B1.
PN
XX 25-SEP-2001.
PD
XX 19-AUG-1998; 98US-00136574.
PF
XX 19-SEP-1997; 97US-00932571.
XX (CLRN ) CLARIANT FINANCE BVI LTD.
PA
PI Farrington GK, Anderson P, Bergquist P, Daniels R, Gibbs MD;
PI Morgan H, Williams DP;
XX
XX WPI; 2002-081780/11.
DR P-PSDB; AAE16325.
XX
XX New cellulase active protein, useful in textile processing or commercial
PT detergents, e.g. for improving the feel or appearance of cotton-
PT containing fabrics, is stable under conditions of alkaline pH and
PT elevated temperatures.
XX
XX Disclosure; Col 71-74; 61pp; English.
XX
CC The present invention relates to a cellulase active protein, which is
CC substantially free of proteinases of native thermophilic and
CC alkaliphilic origin, where the cellulase active protein consists of the
CC CelB5 amino acid sequence. The cellulase active protein is useful for
CC treating cellulosic materials including cotton-containing fabrics, as
CC detergent additives. The cellulase active protein is also useful for
CC improving the feel and/or appearance of cotton-containing fabrics, for
CC removing surface fibers from cotton-containing knits or for imparting
CC stonewashed appearance to cotton-containing denims. The present proteins
CC are stable under condition of alkaline pH and elevated temperatures, thus
CC suitable for textile processing and in commercial detergents. The present
CC sequence is E3/B5 hybrid protein DNA
XX
SQ Sequence 2029 BP; 650 A; 329 C; 520 G; 530 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 4.59e-14 Length: 2029
Score: 195.50 Matches: 38
Percent Similarity: 60.23% Conservative: 15
Best Local Similarity: 43.18% Mismatches: 34
Query Match: 41.68% Indels: 1
DB: 6 Gaps: 1

US-09-917-376-5 (1-88) x AAD26568 (1-2029)
QY 1 ValSerGlyGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
DB 1 ATGGGAGTGGTGTGAAGGTACTGTACAAGAACAAATGACACAAAGTGGCAGCAGGTTCT 60
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrVal 40
DB 61 ATAAGCCGCGTGTGTAAAGATAGTGAATGGAGCGCAGCAGTGTGTGATCTTAGCAGGTT 120
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60

```



Db	121	AAAGATAAGTACTGGTACACAGTGGTGCACAGCCACAGAGTCCGGTA---TGTGAC	177
QY	61	TpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr	80
Db	178	TGGGCACAGATAGGGGCAAGCAATGTGACATTTCAATTTTGTGAAGCTTAGCGCGGAGTG	237
QY	81	ProThrAlaAspThrTyrLeuGln	88
Db	238	AGTGAGCGGATATTACCTGGAG	261
RESULT 13			
AAX55662			
ID	AAX55662	standard; DNA; 6415 BP.	
XX			
AC	AAX55662;		
XX			
DT	30-JUL-1999	(first entry)	
XX			
DE	DNA sequence encoding truncated cellulases.		
XX			
KW	Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;		
KW	Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;		
KW	cotton-containing fabric; stonewashing; ss.		
XX			
OS	Unidentified.		
XX			
PN	EP921188-A2.		
XX			
PD	09-JUN-1999.		
XX			
PF	15-SEP-1998;	98EP-00810919.	
XX			
PR	19-SEP-1997;	97US-00932571.	
XX			
PA	(CLRN ) CLARIANT FINANCE BVI LTD.		
PI	Anderson P, Bergquist PL, Daniels RM, Farrington GK, Gibbs MD;		
PI	Morgan H, Williams DP;		
XX			
DR	WPI; 1999-315403/27.		
DR	P-PSDB; AAY13493.		
XX			
PT	New truncated cellulase proteins, useful in detergents and for producing		
PT	'stonewashed' denim.		
XX			
PS	Disclosure; Page 24-25; 65pp; English.		
XX			
CC	The invention relates to a recombinant cellulase active protein free of		
CC	proteinases of native thermophilic and alkaliphilic origin, comprising		
CC	the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,		
CC	Cel 6 or Cel E3/B5, or a stability region from one of the defined full-		
CC	length sequences, or functional equivalents. Cel B5 extends from amino		
CC	acid A101 to P1424 or K1425 or N1426, and Cel B4/5 extends from amino		
CC	acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends from		
CC	amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel E1/2/3		
CC	extends from Y39 to G812, Cel E6 extends from amino acid V1233 to K1751		
CC	and the stability region extends from amino acid E482 to G635 in the		
CC	sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new		
CC	enzymes are useful in laundry detergent compositions to prevent or remove		
CC	staining, backstaining or graying, for use on cellulosic materials		
CC	including cotton-containing fabrics. They are especially useful for		
CC	preventing redeposition of colorant during stonewashing, and for		
CC	processing of textiles where cellulose breakdown is required. The new		
CC	truncated enzymes show reduced redeposition of dye compared to using non-		
CC	truncated cellulase compositions		
XX			
SQ	Sequence 6415 BP; 2067 A; 1082 C; 1689 G; 1541 T; 0 U; 36 Other;		
Alignment Scores:			
Fred. No.:	2.66e-13	Length:	6415
Score:	194.50	Matches:	38
Percent Similarity:	61.18%	Conservative:	14

Best Local Similarity:	44.71%	Mismatches:	32
Query Match:	41.47%	Indels:	1
DB:	2	Gaps:	1
US-09-917-376-5 (1-88) x AAX55662 (1-6415)			
QY	4	GlyValIysValGlnTyrIlyAsnIleArgAlaSerPheGlyAspAlaProGlyAspAsnGlnIleLysPro	23
Db	2664	GGTGTGAAGGTACTGTACAGAACAATGACAGTGCAGACAGTCTTATAGGCCG	2723
QY	24	GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg	43
Db	2724	TGGTTTAAGATAGTGAATGGAGCGACGACAGTGTGTGATCTTAGCAGGGTTAAGATAAGA	2783
QY	44	TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla	63
Db	2784	TACTGTACACAGTGTGATGTCACAGCCACAGAGTCCGGTA---TGTGACTGGGCACAG	2840
QY	64	MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla	83
Db	2841	ATAGGGCAAGCAATGTGACATTTTGTGAAGCTTAGCAGCGAGTGTGGAGCG	2900
QY	84	AspThrTyrLeuGln	88
Db	2901	GATTATTACCTGGAG	2915
RESULT 14			
AAD26526			
ID	AAD26526	standard; DNA; 6416 BP.	
XX			
AC	AAD26526;		
XX			
DT	26-MAR-2002	(first entry)	
XX			
DE	Active cellulase protein, celE gene.		
XX			
KW	Active cellulase protein; alkaliphilic; textile processing; proteinase;		
KW	detergent additive; stonewashed appearance; cotton-containing denim;		
KW	CelB5; thermophilic; commercial detergent; celE gene; ds.		
XX			
OS	Unidentified.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	634..5889	
FT		/*tag= a	"CelE protein"
FT		/product=	748..2538
FT	misc_feature	/*tag= c	
FT		/product=	"DNA encoding E1/2 protein"
FT	misc_feature	748..2076	
FT		/*tag= b	
FT		/product=	"DNA encoding E1 protein"
PN	US6294366-B1.		
XX			
PD	25-SEP-2001.		
XX			
PF	19-AUG-1998;	98US-00136574.	
XX			
PR	19-SEP-1997;	97US-00932571.	
XX			
PA	(CLRN ) CLARIANT FINANCE BVI LTD.		
PI	Farrington GK, Anderson P, Bergquist P, Daniels R, Gibbs MD;		
PI	Morgan H, Williams DP;		
XX			
DR	WPI; 2002-081780/11.		
DR	P-PSDB; AAE16324.		
XX			
PT	New cellulase active protein, useful in textile processing or commercial		
PT	detergents, e.g. for improving the feel or appearance of cotton-		
PT	containing fabrics, is stable under conditions of alkaline pH and		
PT	elevated temperatures.		



XX PS Disclosure; Col 37-44; 61pp; English.

XX CC The present invention relates to a cellulase active protein, which is

XX CC substantially free of proteinases of native thermophilic and

XX CC alkalophilic origin, where the cellulase active protein consists of the

XX CC CelB5 amino acid sequence. The cellulase active protein is useful for

XX CC treating cellulosic materials including cotton-containing fabrics, as

XX CC detergent additives. The cellulase active protein is also useful for

XX CC improving the feel and/or appearance of cotton-containing fabrics, for

XX CC removing surface fibers from cotton-containing knits or for imparting

XX CC stonewashed appearance to cotton-containing denims. The present proteins

XX CC are stable under condition of alkaline pH and elevated temperatures, thus

XX CC suitable for textile processing and in commercial detergents. The present

XX CC sequence is celE gene

XX SQ Sequence 6416 BP; 2068 A; 1082 C; 1689 G; 1541 T; 0 U; 36 Other;

Alignment Scores:

Pred. No.:	2.66e-13	Length:	6416
Score:	194.50	Matches:	38
Percent Similarity:	61.18%	Conservative:	14
Best Local Similarity:	44.71%	Mismatches:	32
Query Match:	41.47%	Indels:	1
DB:	6	Gaps:	1

US-09-917-376-5 (1-88) x AAD26526 (1-6416)

QY 4 GlyVallyValGlnTyrIlyAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23

DB 2665 GGTGTGAAGGTACTGTACAGAACAAATGAGACAGTGGCAGCACAGGTTCTATAAGCGC 2724

QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43

DB 2725 TGGTTTAAGATAGTGAATGGAGCGCAGCAGTGTGTGATCTTAGCAGGGTTAAGATAAGA 2784

QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCyaspTTPAlaAla 63

DB 2785 TACTGGTACACAGTGGTGGTGCAGCAGCCACAGTGGCGTA---TGTGACTGGGCACAG 2841

QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83

DB 2842 ATAGGGGCAAGCAATGTGACATTCAATTTTGTGAAGCTTAGCAGCGGAGTGAGTGAGCG 2901

QY 84 AspThrTyrLeuGln 88

DB 2902 GATTATTACCTGGAG 2916

RESULT 15

AXX55661

ID AAX55661 standard; DNA; 11706 BP.

XX AC AAX55661;

XX DT 30-JUL-1999 (first entry)

XX DE DNA sequence encoding truncated cellulases Cel B4/5 and Cel B5.

XX KW Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;

XX KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;

XX KW cotton-containing fabric; stonewashing; ss.

XX OS Unidentified.

XX PN EP921188-A2.

XX PD 09-JUN-1999.

XX PF 15-SEP-1998; 98EP-00810919.

XX PR 19-SEP-1997; 97US-00932571.

XX PA (CLRN ) CLARIANT FINANCE BVI LTD.

XX PI Anderson P, Bergquist PL, Daniels RM, Farrington GK, Gibbs MD;

XX DR Morgan H, Williams DP;

XX WPI; 1999-315403/27.

XX P-P8DB; AAY13492.

XX PT New truncated cellulase proteins, useful in detergents and for producing

XX PT 'stonewashed' denim.

XX PS Disclosure; Page 20-23; 65pp; English.

XX CC The invention relates to a recombinant cellulase active protein free of

XX CC proteinases of native thermophilic and alkalophilic origin, comprising

XX CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,

XX CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-

XX CC length sequences, or functional equivalents. Cel B5 extends from amino

XX CC acid A1011 to P1424 or K1425 or N1426, and Cel B4/5 extends from amino

XX CC acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends from

XX CC amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel E1/2/3

XX CC extends from Y39 to G812, Cel E6 extends from amino acid V1233 to K1751

XX CC and the stability region extends from amino acid E482 to G635 in the

XX CC sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new

XX CC enzymes are useful in laundry detergent compositions to prevent or remove

XX CC staining, backstaining or graying, for use on cellulosic materials

XX CC including cotton-containing fabrics. They are especially useful for

XX CC preventing redeposition of colorant during stonewashing, and for

XX CC processing of textiles where cellulose breakdown is required. The new

XX CC truncated enzymes show reduced redeposition of dye compared to using non-

XX CC truncated cellulase compositions

XX SQ Sequence 11706 BP; 3828 A; 1994 C; 2994 G; 2890 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.02e-12	Length:	11706
Score:	192.50	Matches:	37
Percent Similarity:	62.35%	Conservative:	16
Best Local Similarity:	43.53%	Mismatches:	31
Query Match:	41.04%	Indels:	1
DB:	2	Gaps:	1

US-09-917-376-5 (1-88) x AAX55661 (1-11706)

QY 4 GlyVallyValGlnTyrIlyAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23

DB 4038 GGTGTGAAGGTACTGTACAGAACAAATGAGACAGTGGCAGCACAGGTTCTATAAGCGCG 4097

QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43

DB 4098 TGGTTTAAGATAGTGAATGGAGCGCAGCAGTGTGTGATCTTAGCAGGGTTAAGATAAGA 4157

QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCyaspTTPAlaAla 63

DB 4158 TACTGGTACACAGTGGTGGTGCAGCAGCCACAGTGGCGTA---TGTGACTGGGCACAG 4214

QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83

DB 4215 ATAGGGGCAAGCAATGTGACATTCAATTTTGTGAAGCTTAGCAGCGGAGTGAGTGAGCGC 4274

QY 84 AspThrTyrLeuGln 88

DB 4275 GATTATTACCTGGAG 4289

Search completed: October 5, 2005, 09:21:00

Job time : 142.381 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 5, 2005, 07:43:21 ; Search time 846.19 Seconds  
(without alignments)  
3958.515 Million cell updates/sec

Title: US-09-917-376-5  
Perfect score: 469  
Sequence: 1 VSGGVKQYKKNDSAPGDNQ.....IRASFGSVNPATPTADTVLQ 88

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues  
Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters: -MODEL=frame+p2n.model -DEV=xlh  
-O=/cgn2.1/USPTO.spool/US09917376/runat\_04102005.164328.26911/app.query.fasta\_1.2524  
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09917376 @CGN 1 1 12009 @runat\_04102005.164328.26911 -NCPU=6 -ICPU=3  
-NO MWAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_hic:\*  
4: gb\_est3:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_gss1:\*  
9: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	81	17.3	584	CF480372	CF480372 POL1_65_H
2	79.5	17.0	754	BQ047524	BQ047524 EST596642
3	78.5	16.7	768	CB293085	CB293085 UCRCS01_0
4	77	16.4	906	BQ722991	BQ722991 AGENCOURT
5	76	16.2	376	1 AA043716	AA043716 zK61f10.s
6	76	16.2	606	2 BE362891	BE362891 DGI_90_C1
7	76	16.2	641	6 CD226529	CD226529 CCI_46_E
8	76	16.2	657	2 BE362820	BE362820 DGI_89_C0
9	76	16.2	676	7 CF487649	CF487649 POL1_45_C

10	76	16.2	691	4	BG739392	BG739392 EMI_82_D0
11	76	16.2	724	8	BZ616323	BZ616323 1958e08.g
12	76	16.2	814	4	BI218457	BI218457 602937951
13	76	16.2	905	8	BZ545807	BZ545807 OGALR72TM
14	76	16.2	934	9	CG270741	CG270741 OGWB133TV
15	76	16.2	1001	9	CG270748	CG270748 UI-H-BW1-
16	75.5	16.1	657	2	BF513128	BF513128 UI-H-BW1-
17	75.5	16.1	1041	9	CL974250	CL974250 OBIFFC025
18	75.5	16.1	1058	7	CK231562	CK231562 ILJUMIGEN
19	75.5	16.1	2091	3	HSM800308	AL050295 Homo sapi
20	75	16.0	706	7	CK352071	CK352071 hgfh47H
21	74.5	15.9	465	2	AW491464	AW491464 UI-M-BH3-
22	74.5	15.9	613	6	CA595709	CA595709 wpaic.pko
23	74.5	15.9	650	8	BZ894843	BZ894843 Hg4_0131
24	74	15.8	302	7	CF076729	CF076729 QHK12122.
25	74	15.8	463	9	BX959940	BX959940 Reverse s
26	74	15.8	653	4	BI664037	BI664037 603289338
27	73.5	15.7	261	2	BF875908	BF875908 RC3-ET013
28	73.5	15.7	752	7	CN187748	CN187748 UCRCS05_0
29	73.5	15.7	798	7	CN187749	CN187749 UCRCS05_0
30	73.5	15.7	816	6	CB293980	CB293980 UCRCS01_0
31	73.5	15.7	819	6	CB293984	CB293984 UCRCS01_0
32	73.5	15.7	894	7	CK938508	CK938508 CGF100444
33	73.5	15.7	906	3	AY539888	AY539888 Rattus no
34	73.5	15.7	1039	4	BG334904	BG334904 602461219
35	73.5	15.7	1237	5	BQ065850	BQ065850 AGENCOURT
36	73	15.6	357	7	CF078232	CF078232 QHK1J23.Y
37	73	15.6	360	6	C64738	C64738 C64738 Yuji
38	73	15.6	593	7	CF622108	CF622108 laf13D06.
39	73	15.6	776	9	BX209426	BX209426 Danio rer
40	73	15.6	861	8	BH319803	BH319803 CH230-115
41	73	15.6	894	9	CG077853	CG077853 PUFKJ89TB
42	73	15.6	924	9	CG077856	CG077856 PUFKJ89TD
43	72.5	15.5	380	4	BG816375	BG816375 dad35B08.
44	72.5	15.5	450	1	AL799805	AL799805 AL799805
45	72.5	15.5	533	2	AW963120	AW963120 EST375193

ALIGNMENTS

RESULT 1  
CF480372  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

CF480372 584 bp mRNA linear EST 08-SEP-2003  
POL1\_65\_H07\_b1\_A002 Pollen Sorghum bicolor cDNA clone  
POL1\_65\_H07\_A002 3', mRNA sequence.  
CF480372  
CF480372.1 GI:34509241  
EST.  
Sorghum bicolor (sorghum)  
Sorghum bicolor  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Sorghum.  
1 (bases 1 to 584)  
Cordonnier-Pratt,M.-M., Suzuki,Y., Sugano,S., Klein,R.R., Liang,C.,  
Sun,F., Sullivan,R., Eastman,A., Cannon,R., Kern,B., Morgan,J.,  
Lucas,A., Al-Sheikh,A., Jones,V., Adibi,N., Owen,A., Gao,J. and  
Pratt,L.H.  
EST database from Sorghum: pollen  
Unpublished (2003)  
Other ESTs: POL1\_65\_H07\_g1\_A002  
Contact: Cordonnier-Pratt MM  
Laboratory for Genomics and Bioinformatics  
The University of Georgia, Department of Plant Biology  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 583 0210  
Email: mmpratt@uga.edu

Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in  
the Human Genome Center, University of Tokyo Institute of Medical  
Science; plant material and RNA prepared at Texas A & M University;  
sequencing done in the Laboratory for Genomics and Bioinformatics,  
University of Georgia. Sequence ends have been trimmed to exclude



vector and regions below Phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA.  
Seq primer: Sug3-14 (TAGTCTAGCGCGCGGACG)  
POLYA=yes.

FEATURES  
source

Location/Qualifiers

1..584  
/organism="Sorghum bicolor"  
/mol\_type="mRNA"  
/cultivar="BTx623"  
/db\_xref="taxon:4558"  
/clone="POL1.65.H07.A002"  
/lab\_host="DH10B-T1 phage-resistant E. coli"  
/clone\_lib="Pollen"  
/note="Organ: Pollen; Vector: pME18S-FL3; Site\_1: XhoI; Site\_2: XhoI; The library was prepared from polyA+ RNA from pollen at the late vacuolated-vacuolated stage of development. Pollen was harvested from greenhouse-grown panicles of sorghum line BTx623. Panicles were removed from the flag leaf prior to emergence, when no detectable amylase is present in pollen of male-fertile lines. This stage represents pollen collected from anthers about 8-14 days prior to anthesis. Double-stranded cDNA was cloned unidirectionally into different DraIII sites of the pME18S-FL3 vector (5-prime DraIII site is CACTGTGTG, 3-prime DraIII site is CACCATGTG). XhoI excises the cDNA insert."

ORIGIN

Alignment Scores:  
Pred. No.: 10.5 Length: 584  
Score: 81.00 Matches: 24  
Percent Similarity: 45.35% Conservative: 15  
Best Local Similarity: 27.91% Mismatches: 37  
Query Match: 17.27% Indels: 10  
DB: 7 Gaps: 4

US-09-917-376-5 (1-88) x CF480372 (1-584)

Qy 5 ValLysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGlnLeuLysProGly 24  
Db 45 ATAGCAATCAGCGGAACTCAATGAAATCCGTCGGGAGGAGCGTGTGTCGGCGCGGC 104  
Qy 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg--- 43  
Db 105 -----CGCGCGCGGACAGAACCCCTGCAAACTTCCAAACGATCGGTGCAAGGAAA 152  
Qy 44 ---TyrTrp---PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61  
Db 153 CCATGCTGGGATCTCGCAGGAGGAGGTTTGACACCATGGATATATATATGATGAA 212  
Qy 62 AlaAlaMetGlyCysGlyAsnLeu-----ArgAlaSerPheGlySerValAsnPro 78  
Db 213 GCTTGTGTGAAGTGTGGTGTGCTTTTGGAGACGGTGTCTGTCTGTGTGGTCCG 272  
Qy 79 AlaThrProThrAlaAsp 84  
Db 273 AGCTTGCAGTCTGGTGTAT 290

RESULT 2  
BQ047524  
LOCUS BQ047524 754 bp mRNA linear EST 10-MAR-2003  
DEFINITION EST596642 P. infestans-challenged potato leaf, incompatible reaction Solanum tuberosum cDNA clone BPL117P18 5' end, mRNA sequence.

ACCESSION BQ047524  
VERSION BQ047524.1 GI:19821510  
KEYWORDS Solanum tuberosum (potato)  
SOURCE Solanum tuberosum  
ORGANISM Solanum tuberosum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; lamids; Solanales; Solanaceae; Solanum.

REFERENCE  
AUTHORS

1 (bases 1 to 754)  
Zhang, P., Hernandez, M., Tornqvist, C.-E., Wirtz, U., Loukoianov, A., Rangel, P., Haberland, G. T., Karamycheva, S. A., Tsai, J., Chiemingo, A., Bougri, O., Buell, C. R., Ronning, C. M., Helgeson, J. and Baker, B. K. Generation of ESTs from Potato Leaves Challenged with Phytophthora infestans, incompatible interaction (2002)  
Unpublished (2002)  
Contact: Robin Buell  
The Institute for Genomic Research  
9712 Medical Center Dr, Rockville, MD 20850, USA  
Email: potato-array@tigr.org  
This clone can be obtained from the University of Arizona Genomics Institute. Orders can be made through URL:  
http://genome.arizona.edu/orders/  
Seq primer: T3

JOURNAL  
COMMENT

FEATURES  
source

Location/Qualifiers

1..754  
/organism="Solanum tuberosum"  
/mol\_type="mRNA"  
/cultivar="Kennebec"  
/db\_xref="taxon:4113"  
/clone="BPL117P18"  
/tissue\_type="leaf"  
/dev\_stage="6 week old"  
/lab\_host="SOLR"  
/clone\_lib="P. infestans-challenged potato leaf, incompatible reaction"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2: XhoI; supplier: UC Berkeley, PGEC; sequencing: The Institute for Genomic Research. Whole plants were challenged with 450,000 sporangia/ml P. infestans isolate US-1 (US940501) in Biotron (Madison, Wisconsin). Leaf tissue was collected at 1, 2, 5, 12, and 24 hours post-challenge and frozen in liquid nitrogen immediately upon removal. Kennebec plants showed no signs of HR. Katahdin plants (susceptible to P. infestans US-1) were used as controls and showed infection. NOTE: We cannot exclude the possibility that this sequence is actually derived from Phytophthora rather than potato."

ORIGIN

Alignment Scores:  
Pred. No.: 22.5 Length: 754  
Score: 79.50 Matches: 29  
Percent Similarity: 38.05% Conservative: 14  
Best Local Similarity: 25.66% Mismatches: 33  
Query Match: 16.95% Indels: 37  
DB: 5 Gaps: 4

US-09-917-376-5 (1-88) x BQ047524 (1-754)

Qy 7 ValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGlnLeuLysProGlyLeuGln 26  
Db 152 GTGCTTATCAAGATCCAGGAAGT---CGGGTAGTAATCTAATATCTCTCTGTTCA 208  
Qy 27 LeuValAsnThrGlySerSerSer-----ValAspLeuSerThrValThrValArgTyrTrpPheThr 34  
Db 209 GTATCAAAATCTGGCACCTCTCCCTTTCCCTGGAAAATGCCCTATATAATGAGTTTCGT 268  
Qy 35 -----ValAspLeuSerThrValThrValArgTyrTrpPheThr 47  
Db 269 AAGGGGAGCCCTCCGAAGTTTCTTGGTATGAAACATTTCTCCACTCGCAATGGGTTCA 328  
Qy 48 ArgAspGly-----GlySerSerThr 54  
Db 329 AGGTTGGCTCAGGATCATTGACACCAAGTGGCTGGGGCTCTAGGCTAGTCTGGAAC 388  
Qy 55 LeuValTyrAsnCysAspTrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGly 74  
Db 389 CTGACCCCAAAATGTTGGGATTTCAAGGCTAGGTTCTGGTACTGTGACTCCAAATGGTGG 448  
Qy 75 SerValAsnProAlaThrProThrAlaAspThrTyrLeu 87



449 GAA-----CCTCCTTCCGAGATAGTTACCTT 475

Db

RESULT 3

CB293085

LOCUS

DEFINITION

albedo cDNA library Citrus sinensis cDNA clone UCRCS01\_05cb12, mRNA sequence.

CB293085

EST.

CB293085.1 GI:28618542

Source

Citrus sinensis

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Sapindales; Rutaceae; Citrus.

REFERENCE

1 (bases 1 to 768)

AUTHORS

Close,T.J., Collura,K., Fenton,R.D., Feuerbacher,O., Kim,H.R., Kudrna,D., Wananaker,S., Wing,R. and Yu,Y.

TITLE

Development of EST Resources and New Genetic Markers for California Citrus

JOURNAL

Unpublished (2003)

COMMENT

Contact: Timothy Close

Department of Botany & Plant Sciences, University of California Riverside, CA, 92521-0124

Tel: 9097873318

Fax: 9097874437

Email: timothy.close@ucr.edu

Seq primer: T3.

FEATURES

source

1..768

/organism="Citrus sinensis"

/mol\_type="mRNA"

/cultivar="Washington navel"

/db\_xref="taxon:2711"

/clone="UCRCS01\_05cb12"

/tissue\_type="kind containing flavedo and albedo"

/dev\_stage="Mature fruit"

/lab\_host="E. coli TJC121"

/clone\_lib="Washington Navel orange cold acclimated flavedo & albedo cDNA library"

/note="Vector: Lambda Uni-ZAP XR, excised phagenid; Site 1: EcoRI; Site 2: XhoI; Plants were grown in the field at University of California, Riverside Agricultural Operations since 1983. The scion was Washington Navel orange and the rootstock Carizzo Citrange. Tissue from mature fruit was harvested at mid-day in January 2002 during a cold spell, when pre-dawn temperatures were approximately -2 to 2 degree C. Approximately 2 cm median sections of the rind were excised in the field from several fruits, then wrapped in aluminum foil and frozen quickly in dry ice. Total RNA was extracted using a phenol extraction procedure described in J. Japanese Soc. Hort. Sci. (1996) 64:809-814. Poly(A) RNA was purified, a cDNA library was made, and 1 million primary lambda cDNA clones were in vivo excised to give a population of pBluescript SK(-) phagenids. All steps to this point were performed in the TJ Close lab at the University of California, Riverside (Fenton). Phagenids were plated, plasmid DNA purified, cDNA clones archived, and DNA sequences determined bi-directionally using an ABI3530 at the Arizona Genomics Institute, University of Arizona (Collura, Feuerbacher, Kim, Kudrna, Wing, Yu). Chromatogram files were transmitted to UC Riverside (by Yu), then processed at UC Riverside (by Wananaker) using the HarVest pipeline (http://harvest.ucr.edu) to remove vector and cloning oligo sequences and various contaminants, and to trim to a high quality region. Sequences that retained a phred 17 region of at least 100 bases were deposited to GenBank."

ORIGIN

Alignment Scores:

Pred. No.: 58.7 Length: 906

US-09-917-376-5 (1-88) x CB293085 (1-768)

Qy

3 GlyGlyValLysValGlnTyrLysAsnAsnSerAlaProGlyAspAsnGlnIleLys 22

Db

612 GCGCGCATCAAC-----AGNACACACACAGCAGCAATATAAAGTAAT 662

Qy

23 ProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrVal 42

Db

663 AATAATACGACGTTGTACTAGTGTAGCGTAGCTGTGTGCAAGTGGGTGACG--- 719

Qy

43 ArgTyrTrpPheThrArgAspGlyGlySerSerThr 54

Db

720 -----TTTTCGAAGCGGAGGAGCAGT 746

RESULT 4

BQ722991

LOCUS

DEFINITION

AGENCOURT 8103733 Lupski\_sympathetic\_trunk Homo sapiens cDNA clone IMAGE:6191052 5', mRNA sequence.

ACCESSION

BQ722991

VERSION

BQ722991.1 GI:21861888

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 906)

AUTHORS

NIH-MGC http://mgc.nci.nih.gov/.

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: c9apbs-r@mail.nih.gov

Tissue Procurement: Dr. James R. Lupski

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov

Plate: LLAM13590 row: n column: 13

High quality sequence stop: 512.

FEATURES

source

1..906

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:6191052"

/sex="male"

/tissue\_type="sympathetic trunk"

/dev\_stage="adult, 16 yr"

/lab\_host="DH10B"

/clone\_lib="Lupski sympathetic trunk"

/note="Vector: pCMV-SPORT6 (Life Technologies); Site 1: NotI; Site 2: SalI; cDNA made by oligo-dT priming. Directionally cloned using the following adaptors: 5'-TCGACCCACGCGTCCG-3' and 5'-GACTAGTCTTACGTCGAGCGGCCGCCCT(15)-3'. Size selected > 1 kb for average insert length 1.9 kb. This is a primary library, non-amplified. Library constructed by Life technologies and donated by J. Lupski, M.D./Ph.D. (Baylor College of Medicine); available through Life Technologies."

ORIGIN

Alignment Scores:

Pred. No.: 58.7 Length: 906



Score: 77.00 Matches: 16  
 Percent Similarity: 49.15% Conservative: 13  
 Best Local Similarity: 27.12% Mismatches: 26  
 Query Match: 16.42% Indels: 4  
 DB: 5 Gaps: 2

US-09-917-376-5 (1-88) x BQ722991 (1-906)

Qy 7 ValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnLysProGlyLeuGln 26  
 Db 548 CTGGCTATACCAATAAAATCTCCAGGAGGCCACCGTGAGTA-----ATGAAG 598

Qy 27 LeuValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyrTrpPhe 46  
 Db 599 GCTGTAAAGTGGGAATGAAGAATAGTGTGAAGAGCTCTTCCAGCACTACTGTGAC 658

Qy 47 ThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAlaMetGly 65  
 Db 659 TCCCGG---GGCGGCGATCGGCCAGCTCTTACACCTGTCATCTGGGCGAGTGTGGA 712

RESULT 5  
 AA043716

LOCUS AA043716 376 bp mRNA linear EST 11-MAY-1997  
 DEFINITION zk61f10.s1 Soares\_pregnant uterus NBHPU Homo sapiens cDNA clone IMAGE:487339 3' similar to gb:L10240 BASIGIN PRECURSOR (HUMAN); mRNA sequence.

ACCESSION AA043716.1 GI:1521725

VERSION EST.  
 KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.P., Chiapelli, B.,

Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,

Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,

Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,

Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,

Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.,

and Marra, M.

Generation and analysis of 280,000 human expressed sequence tags

Genome Res. 6 (9), 807-828 (1996)

97044478

8889549

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 1400 Std Error: 0.00

Seq primer: -40M13 fwd. from Amersham

High quality sequence stop: 212.

FEATURES

source

1..376

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="GDB:3761053"

/db\_xref="taxon:9606"

/clone="IMAGE:487339"

/sex="female"

/dev\_stage="adult"

/lab\_host="DH10B"

/clone\_lib="Soares\_pregnant uterus NBHPU"

/note="Organ: uterus; Vector: pT73-Pac; Site 1: Not I;

Site 2: Eco RI; 1st strand cDNA was primed with a Not I -

oligo (dT) primer [5'

AACGTGAGAATTCGGCGCCCTTTTTTTTTTTTTTTT 3']

double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT73 vector. Library  
 went through one round of normalization. Library  
 constructed by M. Fatima Bonaldo."

ORIGIN

Alignment Scores:  
 Pred. No.: 25.2 Length: 376  
 Score: 76.00 Matches: 20  
 Percent Similarity: 43.42% Conservative: 13  
 Best Local Similarity: 26.32% Mismatches: 23  
 Query Match: 16.20% Indels: 20  
 DB: 1 Gaps: 3

US-09-917-376-5 (1-88) x AA043716 (1-376)

Qy 17 GlyAspAsnGlnLysProGlyLeuGlnLeuValAsnThrGlySerSerValAsp 36  
 Db 64 GGGGGTTCACGAGGAGCCCTGGT---GCTATAGANAGCCGCCCTCTCCAGGGGTGAG 120  
 Qy 37 LeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerThr----- 54  
 Db 121 CTGGGGAGGCTT---CTGCGGTCTGGAGTCCCGCGATGGCGGCGAGTCCCGAGCAAC 177

Qy 55 -----LeuValTyrAsn 58  
 Db 178 CCCCTCCAGAGCTGCCCGGATGCACAGCAAGGAGGGGCTTGGGAGTGACTTGAGGC 237

Qy 59 CysAspTrpAlaAlaMetGlyCysGlyAsnLysArgAlaSerPheGly 74  
 Db 238 TGTACGGGGTGGCCCTCGGGTGTGGCAAGTGAAGTCTCTGTGGGC 285

RESULT 6

BE362891

LOCUS

DEFINITION

sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

clade; Panicoideae; Andropogoneae; Sorghum.

1 (bases 1 to 606)

Cordonnier-Pratt, M.-M., Gingle, A., Marsala, C., Sudman, M. and

Pratt, L.H.

An EST database from Sorghum: dark-grown seedlings

On Jul 20, 2000 this sequence version replaced gi:9304448.

Unpublished (2000)

Contact: Cordonnier-Pratt MM

Laboratory for Genomics and Bioinformatics

The University of Georgia, Department of Plant Biology

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 583 0210

Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions

below Phred quality 16. The threshold for highest quality sequence

is 20. Three-prime sequences, which are obtained with PolyTMix or

T7 sequencing primer, are presented as the reverse complement.

Seq primer: PolyTMix

High quality sequence start: 15

High quality sequence stop: 594

POLYA=No.

Location/Qualifiers

1..606

/organism="Sorghum bicolor"

/mol\_type="mRNA"

/db\_xref="taxon:4558"

/clone\_lib="Dark Grown 1 (DGL)"

/note="Organ: 5-day-old dark-grown seedlings; Vector:

Lambda Zap; Site\_1: XhoI; Site\_2: EcoRI; The library was



made from poly-A RNA in the cloning vector lambda ZAP II.  
Clones to be sequenced were prepared by mass excision."

ORIGIN

Alignment Scores:  
Pred. No.: 46.7 Length: 606  
Score: 76.00 Matches: 23  
Percent Similarity: 44.15% Conservative: 15  
Best Local Similarity: 26.74% Mismatches: 38  
Query Match: 16.20% Indels: 10  
DB: 2 Gaps: 4

US-09-917-376-5 (1-88) x BE362891 (1-606)

QY 5 ValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnLleLysProGly 24  
DB 208 ATAGCAATCAGCGAATCAATCAATCGGTCGGGAGGAGCCTGTGCTGGCCGGGC 267  
QY 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg--- 43  
DB 268 -----CGGCCGGGACAGAACCTCGAACTTCCAAACGATCGTGCAGAGAAA 315  
QY 44 ---TyrTrp---PheThrArgAspGlyGlySerThrLeuValThrValAsnCysAspTrp 61  
DB 316 CCATGTTGGGGATCTGCAGGAGGAGGTTTGACGACATGGAATATATATATGATGAA 375  
QY 62 AlaAlaMetGlyCysGlyAsnLle-----ArgAlaSerPheGlySerValAsnPro 78  
DB 376 GCTTGTGTGAAGTGTGTCGACGTGTTCTTTGGAGACGCTGCTTCTGCTGTCGTCG 435  
QY 79 AlaThrProThrAlaAsp 84  
DB 436 AGCTTGACGTCGTGTAT 453

RESULT 7  
CD226529 641 bp mRNA linear EST 21-MAY-2003  
LOCUS CCC1\_46\_E07\_b1\_A007 Callus culture/cell suspension Sorghum bicolor  
DEFINITION cDNA clone CCC1\_46\_E07\_A007 3', mRNA sequence.

ACCESSION CD226529  
VERSION CD226529.1 GI:30969963  
KEYWORDS EST.  
SOURCE Sorghum bicolor (sorghum)  
ORGANISM Sorghum bicolor  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 641)  
AUTHORS Cordonnier-Pratt,M.-M., Wentzel,V., Suzuki,Y., Sugano,S.,  
Klein,R.R., Liang,C., Sun,F., Sullivan,R., Shah,M., Rathore,K.,  
Eastman,A. and Pratt,L.H.  
An EST database from Sorghum: callus culture and cell suspension  
Unpublished (2003)

TITLE Other\_ESTs: CCC1\_46\_E07.g1\_A007  
JOURNAL Contact: Cordonnier-Pratt MM  
COMMENT Laboratory for Genomics and Bioinformatics  
The University of Georgia, Department of Plant Biology  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 583 0210  
Email: mmpratt@uga.edu

Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in  
the Human Genome Center, University of Tokyo Institute of Medical  
Science; plant material and RNA prepared at Texas A & M University;  
sequencing done in the Laboratory for Genomics and Bioinformatics,  
University of Georgia. Sequence ends have been trimmed to exclude  
vector and regions below Phred quality 16. Three-prime sequences  
are presented as their reverse complement and have been trimmed to  
exclude polyA.  
Seq primer: Sug3 (CGACCTGCAGCTCGAGCACA)

POLYA=yes.

Location/Qualifiers

1. .641

FEATURES

source

/organism="Sorghum bicolor"  
/mol\_type="mRNA"  
/cultivar="RTx430"  
/db\_xref="taxon:4558"  
/clone="CCC1\_46\_E07\_A007"  
/lab\_host="DH10B-T1 phage-resistant E. coli"  
/clone\_lib="callus culture/cell suspension"  
/note="Vector: pME18S-FL3; Site 1: XhoI; Site 2: XhoI; The  
library was prepared from a mixture of polyA+ RNA from  
callus culture tissue and cells in suspension culture.  
Double-stranded cDNA was cloned unidirectionally into  
different DraIII sites of the pME18S-FL3 vector (5-prime  
DraIII site is CACGTGTG, 3-prime DraIII site is  
CACCATGTG). XhoI excises the cDNA insert."

ORIGIN

Alignment Scores:  
Pred. No.: 50.2 Length: 641  
Score: 76.00 Matches: 23  
Percent Similarity: 44.19% Conservative: 15  
Best Local Similarity: 26.74% Mismatches: 38  
Query Match: 16.20% Indels: 10  
DB: 6 Gaps: 4

US-09-917-376-5 (1-88) x CD226529 (1-641)

QY 5 ValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnLleLysProGly 24  
DB 149 ATAGCAATCAGCGAATCAATCAATCGGTCGGGAGGAGCCTGTGCTGGCCGGGC 208  
QY 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg--- 43  
DB 209 -----CGGCCGGGACAGAACCTCGAACTTCCAAACGATCGTGCAGAGAAA 256  
QY 44 ---TyrTrp---PheThrArgAspGlyGlySerThrLeuValThrValThrValArg 61  
DB 257 CCATGTTGGGGATCTGCAGGAGGAGGTTTGACGACATGGAATATATATGATGAA 316  
QY 62 AlaAlaMetGlyCysGlyAsnLle-----ArgAlaSerPheGlySerValAsnPro 78  
DB 317 GCTTGTGTGAAGTGTGTCGACGTGTTCTTTGGAGACGCTGCTTCTGCTGTCGTCG 376  
QY 79 AlaThrProThrAlaAsp 84  
DB 377 AGCTTGACGTCGTGTAT 394

RESULT 8

BE362820 657 bp mRNA linear EST 20-JUL-2000  
LOCUS DG1\_B9\_C02.g2\_A002 Dark Grown 1 (DG1) Sorghum bicolor cDNA, mRNA  
DEFINITION sequence.

ACCESSION BE362820  
VERSION BE362820.1 GI:9304377  
KEYWORDS EST.  
SOURCE Sorghum bicolor (sorghum)  
ORGANISM Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 657)

Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and

Pratt,L.H.

An EST database from Sorghum: dark-grown seedlings

Unpublished (2000)

CONTACT: Cordonnier-Pratt MM

Laboratory for Genomics and Bioinformatics

The University of Georgia, Department of Plant Biology

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 583 0210

Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions

below Phred quality 16. The threshold for highest quality sequence



is 20.

Seq primer: PolyTMix  
High quality sequence start: 89  
High quality sequence stop: 651  
POLYA=No.

FEATURES

Location/Qualifiers  
1. .657  
/organism="Sorghum bicolor"  
/mol\_type="mRNA"  
/db\_xref="taxon:4558"  
/clone\_lib="Dark Grown 1 (DGI)"  
/note="Organ: 5-day-old dark-grown seedlings; Vector: Lambda Zap; Site 1: XhoI; Site 2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."

ORIGIN

Alignment Scores:  
Pred. No.: 51.8 Length: 657  
Score: 76.00 Matches: 23  
Percent Similarity: 44.19% Conservative: 15  
Best Local Similarity: 26.74% Mismatches: 38  
Query Match: 16.20% Indels: 10  
DB: 2 Gaps: 4

US-09-917-376-5 (1-88) x BE362820 (1-657)

Qy 5 VallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnLleLysProGly 24  
Db 204 ATAGCAATCAGCGGAACTCAATGAATCCGGTCCGGAGGAGCGCTGTGTCGGCCCGGC 263  
Qy 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg--- 43  
Db 264 -----CGCGCGGAGACAGACCTTGAACCTTCCAAACGATCGGTGCAAGGAA 311  
Qy 44 ---TyrTrp---PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61  
Db 312 CCATGTTGGGGATCTGCAGGAGGAGGTTGAGCAACATGGAATATATATATATGATGAA 371  
Qy 62 AlaAlaMetGlyCysGlyAsnIle-----ArgAlaSerPheGlySerValAsnPro 78  
Db 372 GCTTGTGTGAAGTGTGTGAGTGTGTTTGGAGACGGTCTTCTGCTCTGTGTCGCG 431  
Qy 79 AlaThrProThrAlaAsp 84  
Db 432 AGCTTGCACTCTGGTGAT 449

CF487649 676 bp mRNA linear EST 08-SEP-2003  
POLI\_45\_C08\_b1\_A002 Pollen Sorghum bicolor cDNA clone  
CF487649  
CF487649.1 GI:34516518  
EST.  
Sorghum bicolor (sorghum)

ORGANISM  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Sorghum.  
1 (bases 1 to 676)  
Cordonnier-Pratt,M.-M., Suzuki,Y., Sugano,S., Klein,R.R., Liang,C.,  
Sun,F., Sullivan,R., Eastman,A., Cannon,R., Kern,B., Morgan,J.,  
Lucas,A., Al-Sheikh,A., Jones,V., Adibi,N., Owen,A., Gao,J. and  
Pratt,L.H.  
EST database from Sorghum: pollen

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Unpublished (2003)  
Other\_Ests: POLI\_45\_C08\_g1\_A002  
Contact: Cordonnier-Pratt MM  
Laboratory for Genomics and Bioinformatics  
The University of Georgia, Department of Plant Biology  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860

Fax: 706 583 0210

Email: mmpratt@uga.edu  
Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in  
the Human Genome Center, University of Tokyo Institute of Medical  
Science; plant material and RNA prepared at Texas A & M University;  
sequencing done in the Laboratory for Genomics and Bioinformatics,  
University of Georgia. Sequence ends have been trimmed to exclude  
vector and regions below Phred quality 16. Three-prime sequences  
are presented as their reverse complement and have been trimmed to  
exclude polyA.

Seq primer: Sug3-14 (TAGCTAGCGCCGCGACC)

POLYA=Yes.

FEATURES

Location/Qualifiers  
1. .676  
/organism="Sorghum bicolor"  
/mol\_type="mRNA"  
/cultivar="BTX623"  
/db\_xref="taxon:4558"  
/clone="POLI\_45\_C08\_A002"  
/lab\_host="DH10B-T1 phage-resistant E. coli"  
/clone\_lib="Pollen"  
/note="Organ: Pollen; Vector: pME18S-FL3; Site 1: XhoI;  
Site 2: XhoI; The library was prepared from polyA+ RNA  
from pollen at the late vacuolated-vacuolated stage of  
development. Pollen was harvested from greenhouse-grown  
panicles of sorghum line BTX623. Panicles were removed  
from the flag leaf prior to emergence, when no detectable  
amylase is present in pollen of male-fertile lines. This  
stage represents pollen collected from anthers about 8-14  
days prior to anthesis. Double-stranded cDNA was cloned  
unidirectionally into different DraIII sites of the  
pME18S-FL3 vector (5-prime DraIII site is CACTGTGTG,  
3-prime DraIII site is CACCATGTG). XhoI excises the cDNA  
insert."

ORIGIN

Alignment Scores:  
Pred. No.: 53.7 Length: 676  
Score: 76.00 Matches: 23  
Percent Similarity: 44.19% Conservative: 15  
Best Local Similarity: 26.74% Mismatches: 38  
Query Match: 16.20% Indels: 10  
DB: 7 Gaps: 4

US-09-917-376-5 (1-88) x CF487649 (1-676)

Qy 5 VallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnLleLysProGly 24  
Db 214 ATAGCAATCAGCGGAACTCAATGAATCCGGTCCGGAGGAGCGCTGTGTCGGCCCGGC 273  
Qy 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg--- 43  
Db 274 -----CGCGCGGAGACAGACCTTGAACCTTCCAAACGATCGGTGCAAGGAA 321  
Qy 44 ---TyrTrp---PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61  
Db 322 CCATGTTGGGGATCTGCAGGAGGAGGTTTGGCAACATGGAATATATATATGATGAA 381  
Qy 62 AlaAlaMetGlyCysGlyAsnIle-----ArgAlaSerPheGlySerValAsnPro 78  
Db 382 GCTTGTGTGAAGTGTGTGACGTGTGTTTGGAGACGGTGTCTTCTGCTCTGTGTCGCG 441  
Qy 79 AlaThrProThrAlaAsp 84  
Db 442 AGCTTGCACTCTGGTGAT 459

ORIGIN

Alignment Scores:  
Pred. No.: 53.7 Length: 676  
Score: 76.00 Matches: 23  
Percent Similarity: 44.19% Conservative: 15  
Best Local Similarity: 26.74% Mismatches: 38  
Query Match: 16.20% Indels: 10  
DB: 7 Gaps: 4

US-09-917-376-5 (1-88) x CF487649 (1-676)

Qy 5 VallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnLleLysProGly 24  
Db 214 ATAGCAATCAGCGGAACTCAATGAATCCGGTCCGGAGGAGCGCTGTGTCGGCCCGGC 273  
Qy 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg--- 43  
Db 274 -----CGCGCGGAGACAGACCTTGAACCTTCCAAACGATCGGTGCAAGGAA 321  
Qy 44 ---TyrTrp---PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61  
Db 322 CCATGTTGGGGATCTGCAGGAGGAGGTTTGGCAACATGGAATATATATATGATGAA 381  
Qy 62 AlaAlaMetGlyCysGlyAsnIle-----ArgAlaSerPheGlySerValAsnPro 78  
Db 382 GCTTGTGTGAAGTGTGTGACGTGTGTTTGGAGACGGTGTCTTCTGCTCTGTGTCGCG 441  
Qy 79 AlaThrProThrAlaAsp 84  
Db 442 AGCTTGCACTCTGGTGAT 459

RESULT 10

BG739392  
LOCUS  
DEFINITION  
EM1\_82\_D05\_g1\_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA  
sequence.  
ACCESSION  
BG739392.1 GI:14089081  
VERSION  
BG739392.1



KEYWORDS EST.  
SOURCE Sorghum bicolor (sorghum)  
ORGANISM Sorghum bicolor  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.  
REFERENCE 1 (bases 1 to 691)  
AUTHORS Reid, S.P., Cordonnier-Pratt, M.-W., Gingle, A. and Pratt, L.H.  
TITLE An EST database from Sorghum: developing embryos  
JOURNAL Unpublished (2000)  
COMMENT Contact: Cordonnier-Pratt MM  
Laboratory for Genomics and Bioinformatics  
The University of Georgia, Department of Plant Biology  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 583 0210  
Email: mmpratt@uga.edu  
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.  
Seq primer: PolyTMix  
High quality sequence start: 50  
High quality sequence stop: 691  
POLYA=No.

## FEATURES

Location/Qualifiers  
1..691

/organism="Sorghum bicolor"

/mol\_type="mRNA"

/db\_xref="taxon:4558"

/clone\_lib="Embryo 1 (EM1)"

/note="Organ: Embryos germinated for 24 hr; Vector:

pBluescript II from Lambda Zap II; Site 1: XhoI; Site 2:

EcoRI; The library was made from poly-A RNA in the cloning

vector lambda ZAP II. Clones to be sequenced were

prepared by mass excision."

## ORIGIN

Alignment Scores:  
Pred. No.: 55.3 Length: 691  
Score: 76.00 Matches: 23  
Percent Similarity: 44.1% Conservative: 15  
Best Local Similarity: 26.7% Mismatches: 38  
Query Match: 16.2% Indels: 10  
DB: 4 Gaps: 4

US-09-917-376-5 (1-88) x BG739392 (1-691)

QY 5 VallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnLeuLysProGly 24  
DB 265 ATAGCAATCAGCCGAACTCAATGAATCCGTCGGGAGGAGCCCTGTGCTGGCCGGGC 324  
QY 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg--- 43  
DB 325 -----CGGCGCGGAGAGAACCTGAACCTTCAACGATCGGTGCAAGGAA 372  
QY 44 ---TyrTrp---PheThrArgAspGlyGlySerSerThrLeuValThrAsnCysAspTrp 61  
DB 373 CCATGTGGGGATCTCGAGGAGGAGGAGGTTTGAGCAACATGGAATATATATGATGAA 432  
QY 62 AlaAlaMetGlyCysGlyAsnLeu-----ArgAlaSerPheGlySerValAsnPro 78  
DB 433 GCTTGTGTGAGTGTGTGACGTGTGTTCTTTGGAGACGTTGTCTTGTCTGTGTCGG 492  
QY 79 AlaThrProThrAlaAsp 84  
DB 493 AGCTTGACGTCTGGTGT 510

## RESULT 11

BZ616323/c 724 bp DNA linear GSS 16-JAN-2003  
LOCUS  
DEFINITION i958e08.g1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone  
i958e08 5', genomic survey sequence.  
ACCESSION BZ616323

## VERSION

KEYWORDS  
SOURCE

ORGANISM

## REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BZ616323.1 GI:27766878

GSS.

Zea mays

Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 724)

Rabinowicz, P.D., O'Shaughnessy, A.L., Balija, V., Dedhia, N.,

Katzenburger, F., King, L., Miller, B., Muller, S., Nascimento, L.,

Zatavarn, I., McCombie, W.R. and Martienssen, R.A.

Genomic shotgun sequences from Zea mays (methyl-filtered)

Unpublished (2002)

Contact: W. Richard McCombie

Lita Annenberg Hazen Genome Sequencing Center

Cold Spring Harbor Laboratory

PO Box 100, Cold Spring Harbor, NY 11724, USA

Tel: 516 367 8884

Fax: 516 367 8874

Email: mcombie@cshl.org

Plate: i958 row: e column: 08

Seq primer: -21M3UnivRev

Class: shotgun

High quality sequence stop: 728.

## FEATURES

source

1..724

/organism="Zea mays"

/mol\_type="genomic DNA"

/cultivar="B73"

/db\_xref="taxon:4577"

/clone="ig58e08"

/lab\_host="DH5a"

/clone\_lib="WGS-ZmaysF (DH5a methyl filtered)"

/note="Organ: Immature ears; Site 1: Xba I; Site 2: Xba I;

The vector was digested with XbaI and one nucleotide was

added by fill in the recessive 3' end. The genomic DNA

was nebulized, end repaired, adaptor ligated and size

fractionated using sephadex. The resulting fragments were

between 0.8 and 3 kb and were cloned into the vector (.x/y

reads in M13mp19, .b/g reads in pUC19). The same ligation

was transformed into DH5a."

## ORIGIN

Alignment Scores:  
Pred. No.: 58.7 Length: 724  
Score: 76.00 Matches: 34  
Percent Similarity: 36.28% Conservative: 7  
Best Local Similarity: 30.09% Mismatches: 29  
Query Match: 16.20% Indels: 43  
DB: 6 Gaps: 6

US-09-917-376-5 (1-88) x BZ616323 (1-724)

QY 2 SerGlyValLysValGlnTyrLysAsnAspSer-----AlaPro 16  
DB 587 TCGCGAGCGTCAGATCCCATCAAGAGCGGCAACGACGCTGGCGAGTCGGGCC 528  
QY 17 Gly-----AspAsnGlnIleLysProGlyLeuGlnLeuValAsn----- 29  
DB 527 GGGCGCCAGCTGAGCAGCTGACGCGGAGCCACGACTGATGCTTATCAACCGTGGCCGC 468  
QY 30 Thr-GlySerSerValAsp----- 36  
DB 467 ACGAGGAAGTTCGTCAACTTCCCGGAGATCGTCGGAGCGGTCCAGAGCGGTGGTTCGAG 408  
QY 37 -----LeuSerThrValThrValArgTyrTrpPheThrArgAspGlySerSe 53  
DB 407 GTCATCCGATCGAGCGCGGTAGAGACTCAGCGGTGAGGAGCTTCGCCGACCGCTCGAC 348  
QY 53 rThrLeuValTyrAsnCysAspTrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPh 73  
DB 347 TCGTGGC-----ACGTGCTCATGGGCGCGCACGGG----- 317



```

Qy 73 eGlySerValAsnProAlaThrProThrAlaAspThr 85
Db 316 -----CGGCGCTCACCACTTCTTCTTCC 293

RESULT 12
B1218457 814 bp mRNA linear EST 11-JUL-2001
LOCUS 602937951P1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:5101069 5',
DEFINITION mRNA sequence.
ACCESSION B1218457
VERSION B1218457.1 GI:14671901
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 814)
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM11243 row: f column: 14
High quality sequence stop: 792.
FEATURES
source
1..814
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5101069"
/lab_host="DH10B (TI phase-resistant)"
/clone_lib="NCI_CGAP_L19"
/note="Organ: liver; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

ORIGIN
Alignment Scores:
Pred. No.: 68.3 Length: 814
Score: 76.00 Matches: 28
Percent Similarity: 44.83% Conservative: 11
Best Local Similarity: 32.18% Mismatches: 29
Query Match: 16.20% Indels: 19
DB: 4 Gaps: 4

US-09-917-376-5 (1-88) x B1218457 (1-814)

Qy 2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
Db 33 GCAGGTGGAGCCCGAGTATGGGCGGAACACACAGATCGCCAGGCAATGAGCAG-- 89
Qy 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
Db 90 -----AGTTCCCGAGTGGAGCCTTGAGGACCCCTG--- 116
Qy 42 ValArgTyrTrpPheThrArgAspGlyGlySerThrLeu-----ValTyrAsnCys 59
Db 117 CTCGCTACTACACACAGGCGGCGGCTCTACAGATCCAGGATGACGGCTGT 176
Qy 60 Asp-TrpAlaAlaMetGlyCysGlyAsnIleAraGalaSerPheGlySerValAsnProAl 79
Db 177 GACGTGGGGTCGAGCGGGGCC-----TCCTCCGCGGTACGAGCAGGTGCGCA 224
Qy 79 aThrProThrAlaAspThr 85

RESULT 13
B2545807 905 bp DNA linear GSS 16-DEC-2002
LOCUS OGALR72TM ZM2 0.7 1.5 KB Zea mays genomic clone ZMMBma0089L24,
DEFINITION genomic survey sequence.
ACCESSION B2545807
VERSION B2545807.1 GI:27099699
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 905)
TITLE Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
AUTHORS Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,
Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.
JOURNAL Consortium for Maize Genomics
COMMENT Unpublished (2002)
Other GSSs: OGALR72TC
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.
FEATURES
source
1..905
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBma0089L24"
/clone_lib="ZM2 0.7 1.5 KB"
/note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Alignment Scores:
Pred. No.: 78.3 Length: 905
Score: 76.00 Matches: 34
Percent Similarity: 36.28% Conservative: 7
Best Local Similarity: 30.09% Mismatches: 29
Query Match: 16.20% Indels: 43
DB: 8 Gaps: 6

US-09-917-376-5 (1-88) x B2545807 (1-905)

Qy 2 SerGlyGlyValLysValGlnTyrLysAsnAspSer-----AlaPro 16
Db 87 TCCGCGCGGTGAGCATCCATACAGGAGGCGGCAACGCTCGGCGAGTGGCCC 146
Qy 17 Gly-----AspAsnGlnIleLysProGlyLeuGlnLeuValAsn----- 29
Db 147 GGCGCCCGCAGCTGAGCAGCTGCAGCGGAGGCCACGACTGATCTTATCAACCGTGGCGC 206
Qy 30 Thr-GlySerSerValAsp----- 36
Db 207 ACGAGGAAGTTCTCAACTTCCCGAGATCTCCGAGCGGTCCAGAGCGGTCCAGAGCTGGGTTTCGAG 266
Qy 37 -----LeuSerThrValThrValArgTyrTrpPheThrArgAspGlySerSe 53
Db 267 GTCATCCGATCGAGCCCGGTAGAGACCTCAGCGTGGAGGACTTCGCCCGGACCGTGCAC 326
Qy 53 rThrLeuValTyrAsnCysAspTrpAlaAlaMetGlyCysGlyAsnIleAraGalaSerPh 73
Db 327 TCGTGGC-----ACGTGCTCATGGCGCGCAGGGG----- 357

```



Qy 73 eGlySerValAsnProAlaThrProThrAlaAspThr 85  
 Db 358 -----CCGGCCTCACCACTTCTTCTTCC 381

## RESULT 14

CG270741/c  
 LOCUS CG270741 934 bp DNA linear GSS 25-AUG-2003  
 DEFINITION OGWB33TH ZM\_0.7\_1.5 KB Zea mays genomic clone ZMMBma0583E18,  
 genomic survey sequence.

ACCESSION CG270741  
 VERSION CG270741.1 GI:34182882  
 KEYWORDS GSS.

SOURCE Zea mays  
 ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 934)

AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,  
 Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,  
 Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.

TITLE Consortium for Maize Genomics

## JOURNAL

COMMENT Other GSSs: OGWB33TV

Contact: Cathy Whitelaw

TIGR

7912 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TF

Class: sheared ends.

Location/Qualifiers

## FEATURES

source

1..934  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone="ZMMBma0583E18"  
 /clone\_lib="ZM\_0.7\_1.5 KB"  
 /note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb  
 methylation filtered genomic DNA library"

## ORIGIN

Alignment Scores:  
 Pred. No.: 81.5 Length: 934  
 Score: 76.00 Matches: 34  
 Percent Similarity: 36.28% Conservative: 7  
 Best Local Similarity: 30.09% Mismatches: 29  
 Query Match: 16.20% Indels: 43  
 DB: 9 Gaps: 6

US-09-917-376-5 (1-88) x CG270741 (1-934)

Qy 2 SerGlyValLysValGlnTyrLysAsnAspSer-----AlaPro 16  
 Db 804 TCCGAGGCGTCAGCATCCATACAAAGGAGGCGGACAGCAACGCTGCGGAGTGGCGCC 745

Qy 17 Gly-----AspAsnGlnIleLysProGlyLeuGlnLeuValAsn----- 29

Db 744 GCGGCCCGCAGCTGACGAGCTGACGCGGAGGAGCCACGACTGATGCTTATCAACCGTGGCCGC 685

Qy 30 Thr-GlySerSerSerValAsp----- 36

Db 684 ACAGAGAGTTCGTCAACTTCCGAGAGATCGTCGGAGCGGTCCAGAGGCTGGGTTCGAG 625

Qy 37 -----LeuSerThrValThrValArgTyrTirpPheThrArgAspGlyGlySerSe 53

Db 624 GTCATCCGATGAGCCGCGTAGAGACCTCAGCGTGGAGGACTTCGCCCGGACCGTCGAC 565

Qy 53 rThrLeuValTyrAsnCysAspTirpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPh 73

Db 564 TCGTGG-----ACGTGCTCATGGCGCGCGCGGG----- 534

Qy 73 eGlySerValAsnProAlaThrProThrAlaAspThr 85  
 Db 533 -----CCGGCCTCACCACTTCTTCTTCC 510

## RESULT 15

CG270748  
 LOCUS CG270748 1001 bp DNA linear GSS 25-AUG-2003  
 DEFINITION OGWB33TV ZM\_0.7\_1.5 KB Zea mays genomic clone ZMMBma0583E18,  
 genomic survey sequence.

ACCESSION CG270748  
 VERSION CG270748.1 GI:34182889  
 KEYWORDS GSS.

SOURCE Zea mays

ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 1001)

AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,  
 Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,  
 Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.

TITLE Consortium for Maize Genomics

## JOURNAL

COMMENT Other GSSs: OGWB33TH

Contact: Cathy Whitelaw

TIGR Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TF

Class: sheared ends.

Location/Qualifiers

## FEATURES

source

1..1001  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone="ZMMBma0583E18"  
 /clone\_lib="ZM\_0.7\_1.5 KB"  
 /note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb  
 methylation filtered genomic DNA library"

## ORIGIN

Alignment Scores:  
 Pred. No.: 89.1 Length: 1001  
 Score: 76.00 Matches: 34  
 Percent Similarity: 36.28% Conservative: 7  
 Best Local Similarity: 30.09% Mismatches: 29  
 Query Match: 16.20% Indels: 43  
 DB: 9 Gaps: 6

US-09-917-376-5 (1-88) x CG270748 (1-1001)

Qy 2 SerGlyValLysValGlnTyrLysAsnAspSer-----AlaPro 16  
 Db 418 TCCGAGGCGTCAGCATCCATACAAAGGAGGCGGACAGCAACGCTGCGGAGTGGCGCC 477

Qy 17 Gly-----AspAsnGlnIleLysProGlyLeuGlnLeuValAsn----- 29

Db 478 GCGGCCCGCAGCTGACGAGCTGACGCGGAGGAGCCACGACTGATGCTTATCAACCGTGGCCGC 537

Qy 30 Thr-GlySerSerSerValAsp----- 36

Db 538 ACAGAGAGTTCGTCAACTTCCGAGAGATCGTCGGAGCGGTCCAGAGGCTGGGTTCGAG 597

Qy 37 -----LeuSerThrValThrValArgTyrTirpPheThrArgAspGlyGlySerSe 53

Db 598 GTCATCCGATGAGCCGCGTAGAGACCTCAGCGTGGAGGACTTCGCCCGGACCGTCGAC 657

Qy 53 rThrLeuValTyrAsnCysAspTirpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPh 73



Db 658 TCGTGG-----ACGTGCTCATGGGGCGGCACGGGG----- 688  
Qy 73 eGlySerValAsnProAlaThrProThrAlaAspThr 85  
Db 689 -----CCGGCCTCACCAACTTCTCTTCC 712

Search completed: October 5, 2005, 21:09:34  
Job time : 848.19 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 5, 2005, 07:47:30 ; Search time 41.8869 Seconds  
(without alignments)  
3437.649 Million cell updates/sec

Title: US-09-917-376-5  
Perfect score: 469  
Sequence: 1 VSGGVKQYKXNDSPAGDQ.....IRASFGSVNPATPTADTVLQ 88

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09917376/runat 04102005 164328 26925/app query.fasta\_l.2524  
-DB=Issued Patents NA -QWMT=fastcap -SUPFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09917376 @CGN 1.1 516 @runat 04102005 164328 26925 -NCFU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOB=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/6C.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	195.5	41.7	2029	3	US-09-136-574A-46 Sequence 46, Appl
2	194.5	41.5	6416	3	US-09-136-574A-2 Sequence 2, Appl
3	192.5	41.0	11707	3	US-09-136-574A-1 Sequence 1, Appl
4	182.5	38.9	2977	2	US-07-862-588B-1 Sequence 1, Appl
5	165	35.2	1438	4	US-09-339-159B-3 Sequence 3, Appl
6	165	35.2	1482	3	US-09-198-956-9 Sequence 9, Appl
7	165	35.2	1482	3	US-09-198-955A-11 Sequence 11, Appl
8	165	35.2	1482	3	US-09-694-531-11 Sequence 11, Appl
9	165	35.2	1482	3	US-09-670-141-9 Sequence 9, Appl
10	165	35.2	1482	4	US-10-072-152-11 Sequence 11, Appl
11	155.5	33.2	4059	4	US-09-784-554B-1 Sequence 1, Appl
12	150.5	32.1	4056	4	US-09-784-554B-3 Sequence 3, Appl

13	120.5	25.7	1624	2	US-07-862-588B-6 Sequence 6, Appl
14	120.5	25.7	1775	2	US-07-862-588B-5 Sequence 5, Appl
15	114	24.3	486	1	US-08-048-164A-1 Sequence 1, Appl
16	114	24.3	486	1	US-08-048-164A-3 Sequence 3, Appl
17	114	24.3	486	1	US-08-460-462-1 Sequence 1, Appl
18	114	24.3	486	1	US-08-460-462-3 Sequence 3, Appl
19	114	24.3	486	1	US-08-460-457-1 Sequence 1, Appl
20	114	24.3	486	1	US-08-460-457-3 Sequence 3, Appl
21	114	24.3	486	1	US-08-460-458-1 Sequence 1, Appl
22	114	24.3	486	1	US-08-460-458-3 Sequence 3, Appl
23	114	24.3	486	2	US-08-460-455-1 Sequence 1, Appl
24	114	24.3	486	2	US-08-460-455-3 Sequence 3, Appl
25	114	24.3	486	2	US-08-330-394A-1 Sequence 1, Appl
26	114	24.3	486	2	US-08-330-394A-3 Sequence 3, Appl
27	114	24.3	499	3	US-09-006-636-5 Sequence 5, Appl
28	114	24.3	499	3	US-09-006-632-5 Sequence 5, Appl
29	114	24.3	499	3	US-09-325-274-5 Sequence 5, Appl
30	113	24.1	1146	3	US-09-277-716-21 Sequence 21, Appl
31	113	24.1	1146	3	US-09-609-161B-21 Sequence 21, Appl
32	77	16.4	8107	4	US-09-335-586-3 Sequence 3, Appl
33	76	16.2	136058	4	US-09-949-016-12565 Sequence 12565, A
34	76	16.2	136480	4	US-09-949-016-17064 Sequence 17064, A
35	75.5	16.1	5714	4	US-09-620-312D-393 Sequence 393, App
36	72.5	15.5	819	4	US-09-902-540-3752 Sequence 3752, Ap
37	72.5	15.5	17897	4	US-09-902-540-1182 Sequence 1182, Ap
38	72.5	15.5	20284	4	US-09-526-193A-21 Sequence 21, Appl
39	70	14.9	10892	4	US-09-902-540-962 Sequence 962, App
40	70	14.9	25165	3	US-09-453-702B-39 Sequence 39, Appl
41	70	14.9	246444	4	US-09-949-016-13113 Sequence 13113, A
42	69.5	14.8	103987	4	US-09-949-016-12513 Sequence 12513, A
43	69.5	14.8	103988	4	US-09-949-016-17050 Sequence 17050, A
44	68.5	14.6	601	4	US-09-949-016-128395 Sequence 128395,
45	68.5	14.6	601	4	US-09-949-016-128396 Sequence 128396,

ALIGNMENTS

RESULT 1

US-09-136-574A-46  
; Sequence 46, Application US/09136574A  
; Patent No. 6294366  
; GENERAL INFORMATION:

APPLICANT: Farrington, Graham K.  
Anderson, Paige  
Gibbs, Moreland  
Bergquist, Peter  
Daniels, Roy  
Morgan, Hugh W.  
Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for  
Treating Cellulose Containing Fabrics Using Truncated  
Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49

CORRESPONDENCE ADDRESS:

ADDRESSER: Howson and Howson

STREET: Spring House Corporate Center, P.O. Box 457

CITY: Spring House

STATE: PA

COUNTRY: USA

ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/136,574A

FILING DATE: 19-Aug-1998

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/932,571

FILING DATE: September 19, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Bak, Mary E.



```

/ / REGISTRATION NUMBER: 31,215
/ / REFERENCE/DOCKET NUMBER: 13997US001/CIP
/ / TELECOMMUNICATION INFORMATION:
/ / TELEPHONE: 215-540-9200
/ / TELEFAX: 215-540-5818
/ / TELEX: <Unknown>
/ / INFORMATION FOR SEQ ID NO: 46:
/ / SEQUENCE CHARACTERISTICS:
/ / LENGTH: 2029 base pairs
/ / TYPE: nucleic acid
/ / STRANDEDNESS: single
/ / TOPOLOGY: linear
/ / SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-09-136-574A-46

```

Alignment Scores:	
Pred. No.:	6.46e-17
Score:	195.50
Length:	2029
Matches:	38
Conservative:	15
Mismatches:	34
Best Local Similarity:	43.18%
Query Match:	41.68%
Indels:	1
Gaps:	1
DB:	3

US-09-917-376-5 (1-88) X US-09-136-574A-46 (1-2029)

Qy	1	ValSerGlyGlyValLysValGlnTyrIlyshsnshnAspSerAlaProGlyAspAsnGln	20
Db	1	ATGGGAAGTGCGTGTGAAGGTACTGTACAAGAACAAATGAGCAACAGTCGCGACACAGGTTCT	60
Qy	21	IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal	40
Db	61	ATAGGCCGTGGTTTAAGATAGTGAATGGAGCGACGACAGCTGTTGATCTTAGCAGGTT	120
Qy	41	ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp	60
Db	121	AAAGATAAGATACATCGTACACACATGCGATGGTCACCAAGCCACAGAGTCGCGGTA--TGTGAC	177
Qy	61	TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr	80
Db	178	TGGGCACAGATAGGGCGCAAGCAATGTGCATTCATTTTGTGAAGCTTAGCAGCGGAGTG	237
Qy	81	ProThrAlaAspThrTyrLeuGln	88
Db	238	AGTGAGCGGGATTATTACTCTGGAG	261

RESULT 2  
US-09-136-574A-2

? Patent NO. 6294366  
 ? GENERAL INFORMATION:  
 ? APPLICANT: Farrington, Graham K.  
 ? Anderson, Paige  
 ? Gibbs, Moreland  
 ? Bergquist, Peter  
 ? Daniels, Roy  
 ? Morgan, Hugh W.  
 ? Williams, Diane P.  
 ?

APPLICATION NUMBER: US 09/136,574  
 FILING DATE: 19-AUG-1998  
 CLASSIFICATION: <UNKNOWN>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/932,571  
 FILING DATE: September 19, 1997  
 ATTORNEY/AGENT INFORMATION:

```

, INFORMATION FOR SEQ ID NO: 2:
,   SEQUENCE CHARACTERISTICS:
,     LENGTH: 6416 base pairs
,     TYPE: nucleic acid
,     STRANDEDNESS: single
,     TOPOLOGY: linear
,   SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-136-574A-2

```

Alignment Scores:		
Pred. No.:	4,68e-16	Length: 6416
Score:	194.50	Matches: 38
Percent Similarity:	61.18%	Conservative: 14
Best Local Similarity:	44.71%	Mismatches: 32
Query Match:	41.47%	Indels: 1
DB:	3	Gaps: 1
US-09-917-376-5 (1-88) x US-09-136-574A-2 (1-6416)		

Qy		4	GlyValIysValGlnTyrLysAsnAnaspSerAlapProGlyAspAenGlinLeLysPro	23
Dq			:         :	
Db		2865	GGTGGAAGGTACTGCACAGAACAATGAGCAGATGCCGAGCACAGCTTTATAAGCCCG	2724
			:   :   :   :   :   :	
Qy		24	GlyLeuGlnLeuValAsnThrGlySerSerValasPleusSerThrValThrValArg	43
Dq			:   :   :   :   :   :	
Db		2725	TGGTTTAAGATAGTAGTCGAGGCGAGCAAGCTGTTGATTCTTAGCAGGGTTAAGATAAGA	2784
			:   :   :   :   :   :	
Qy		44	TyrTrpPhethrArgaspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaala	63
Dq			:         :	
Db		2785	TACTGTGTACCACAGTGGATGGTGTGCAAGCCCACAGAGTGGCGGTA--TGTGACTGGGGCACAG	2841
			:         :	
Qy		64	MetGlyCysGlyAsnileArgAlaserPheGlySerValAsnProlathrProthrAla	83
Dq			:   :   :   :   :   :	
Db		2842	ATAGGGCGAACCATGTGCATTCATTTTTGTGAAGCTTAGCAGCGAGTGGAGCG	2901
			:         :	
Qy		84	AspThrTyrLeuGln	88
Dq			:         :	
Db		2902	GATTATTACCTGGAG	2916

RESULT 3  
US-09-136-574A-1  
; Sequence 1, Application US/09136574A  
; Patent No. 6294366

APPLICANT: Farrington, Graham K.  
Anderson, Paige  
Gibbs, Moreland  
Bergquist, Peter  
Daniels, Roy  
Morgan, Hugh W.  
Williams, Diane P.



STATE: PA  
COUNTRY: USA  
ZIP: 19477  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/136,574A  
FILING DATE: 19-Aug-1998  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/932,571  
FILING DATE: September 19, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: 1997US001/CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-540-9200  
TELEFAX: 215-540-5818  
TELEX: <Unknown>  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11707 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-136-574A-1

Alignment Scores:  
Pred. No.: 2,11e-15 Length: 11707  
Score: 192.50 Matches: 37  
Percent Similarity: 62.35% Conservative: 16  
Best Local Similarity: 43.53% Mismatches: 31  
Query Match: 41.04% Indels: 1  
DB: 3 Gaps: 1

US-09-917-376-5 (1-88) x US-09-136-574A-1 (1-11707)

QY 4 GlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnLeuLysPro 23  
Db 4038 GGTGTTGAGGTACTATACAGAACATGACAGGCGAGCAGCAAGTTCTATAGGCGG 4097  
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43  
Db 4098 TGGTTTATAGATAGTGAATGAGGCGACAGCAGTGTGATCTTAGCAGGGTTAAGATAAG 4157  
QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63  
Db 4158 TACTGGTACACAGTGGTGTGACAGCCACAGAGTGGCGTA---TGTGACTGGGACAG 4214  
QY 64 MetGlyCysGlyAsnLeuArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83  
Db 4215 ATAGGGGCAAGCAATGTGACATTCATTTTGTGAGCTGAGCAGCGAGTGAGTGAGCGG 4274  
QY 84 AspThrTyrLeuGln 88  
Db 4275 GATTATTACTTGGAG 4289

RESULT 4

US-07-862-588B-1  
Sequence 1, Application US/07862588B  
Patent No. 5916796  
GENERAL INFORMATION:  
APPLICANT: Joergensen, Per Linnae  
APPLICANT: Sch lein, Martin  
APPLICANT: Hansen, Christian  
TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 5916796 No. 5916796disk of No. 5916796th America, Inc.

STREET: 405 Lexington Avenue, 62nd floor  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10017  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/862,588B  
FILING DATE: 19920727  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DK 164/90  
FILING DATE: 19-JAN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/DK91/00013  
FILING DATE: 18-JAN-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Zelson, Steve T. / Lambiris, Elias J.  
REGISTRATION NUMBER: 30,335 / 33,728  
REFERENCE/DOCKET NUMBER: 3425.204-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 867 0123  
TELEFAX: 212 867 0298  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2977 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Bacillus lautus  
STRAIN: NCIMB 40250  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 677..2776  
OTHER INFORMATION:  
US-07-862-588B-1  
Alignment Scores:  
Pred. No.: 7.14e-15 Length: 2977  
Score: 182.50 Matches: 36  
Percent Similarity: 66.27% Conservative: 19  
Best Local Similarity: 43.37% Mismatches: 25  
Query Match: 38.91% Indels: 3  
DB: 2 Gaps: 2  
US-09-917-376-5 (1-88) x US-07-862-588B-1 (1-2977)  
QY 6 LysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnLeuLysProGlyLeu 25  
Db 144 CGACTCGAGTACAGCGCGCGGATACAAATGCAGCCGACCAACAGATCAAGCCGCTTC 203  
QY 26 GlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyrTrp 45  
Db 204 AACATCAAAACAAACGCTACTTCGGCTGTGATTAAAGCAGCGCTCAAAATCCGCTACTAC 263  
QY 46 PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAlaMetGly 65  
Db 264 TTCACCAAGATGGTTCTCGCGCGGTGACAGCGTGG---ATCGACTGGGCGCAGCTCGGC 320  
QY 66 CysGlyAsnLeuArgAlaSerPheGlySerValAsnProAlaThrProThrAlaAspThr 85  
Db 321 GGCAGCAACATTCAGATCTCTGTTGGC-----AACCATCTGGCAGCAATTCGATAG 374  
QY 86 TyrLeuGln 88  
Db 375 TACGTGGAG 383



```
RESULT 5
US-09-339-159B-3
; Sequence 3, Application US/09339159B
; Patent No. 6566114
; ORGANISM: Bacillus sp. 1633
; GENERAL INFORMATION:
; APPLICANT: Kauppinen, Markus
; APPLICANT: Schulein, Martin
; APPLICANT: Schnorr, Kirk
; APPLICANT: Andersen, Lene
; APPLICANT: Bjornvad, Mads
; TITLE OF INVENTION: No. 6566114el Mannanases
; FILE REFERENCE: 5440.204-US
; CURRENT APPLICATION NUMBER: US/09/339,159B
; CURRENT FILING DATE: 1999-06-24
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1438
; TYPE: DNA
; ORGANISM: Bacillus sp. 1633
US-09-339-159B-3

Alignment Scores:
Pred. No.: 6.68e-13 Length: 1438
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.18% Indels: 10
DB: 4 Gaps: 2

US-09-917-376-5 (1-88) x US-09-339-159B-3 (1-1438)
Qy 1 ValSerGlyGlyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 940 GTATCAGGCAATTGGAAGTTGAAATCTACACGAGCAGTGCATCTTCACTACTCA 999
Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1000 ATCAATCTCTCAGTTCAAGTTACTAATACCGAAGCAGTGCATCTTCACTACTCA 1059
Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1060 ACATTGAGATATTATTATACAGTACGACGAGCAGAAAGATCAGACCTTCTGG--TGTGAC 1116
Qy 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1117 CATGCTGCAATAATCGGCAAGTACGAGCAGTACAAACGGAATTACTTCAAAATGTAAAGGA 1176
Qy 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1177 ACATTGTAAATAATGAGTTCTTCAACAATAACGAGACACTACTCTTGAA 1227

RESULT 6
US-09-198-956-9
; Sequence 9, Application US/09198956
; Patent No. 6165769
; ORGANISM: Bacillus licheniformis
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels Erik K.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Schnorr, Kirk
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
; FILE REFERENCE: 5377.200-US
; CURRENT APPLICATION NUMBER: US/09/198,956
; CURRENT FILING DATE: 1998-11-24
; EARLIER APPLICATION NUMBER: 1344/97
; EARLIER FILING DATE: 1997-11-24
; EARLIER APPLICATION NUMBER: 60/067,240
; EARLIER FILING DATE: 1997-12-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-198-956-9

Alignment Scores:
Pred. No.: 6.97e-13 Length: 1482
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.18% Indels: 10
DB: 3 Gaps: 2

US-09-917-376-5 (1-88) x US-09-198-956-9 (1-1482)
Qy 1 ValSerGlyGlyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGGCAATTGGAAGTTGAAATCTACACGAGCAGTGCATCTTCACTACTCA 1050
Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCTCTCAGTTCAAGTTACTAATACCGAAGCAGTGCATCTTCACTACTCA 1110
Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATTATACAGTACGACGAGCAGAAAGATCAGACCTTCTGG--TGTGAC 1167
Qy 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCAATAATCGGCAAGTACGAGCAGTACAAACGGAATTACTTCAAAATGTAAAGGA 1227
Qy 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTGTAAATAATGAGTTCTTCAACAATAACGAGACACTACTCTTGAA 1278

RESULT 7
US-09-198-955A-11
; Sequence 11, Application US/09198955A
; Patent No. 6187580
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Samne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6187580el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/09/198,955A
; CURRENT FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Clostridium thermocellum
US-09-198-955A-11
```



```
Alignment Scores:
Pred. No.: 6.97e-13 Length: 1482
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservativeness: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.18% Indels: 10
DB: 3 Gaps: 2

US-09-917-376-5 (1-88) x US-09-198-955A-11 (1-1482)
QY 1 ValSerGlyGlyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGCAATTTGAAGTTGAATCTACACAGCAATCTTCAGATACTACTACTCA 1050
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCCTCAGTTCAAGTTTACTAATACCGAGCAGTGCATTTGATTTGTCCAAACTC 1110
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATTATACAGTACGACGAGCAAGAGATCAGACCTTCTGG---TGTGAC 1167
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCAATTAATCGGCAGTAAACGGCAGTACACGGAATTAATCTCAATGTAAAGGA 1227
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTTGTAAATAATGAGTTCTCTCAACAATAACGACAGACCTACCTTGAA 1278

RESULT 8
US-09-694-531-11
; Sequence 11, Application US/09694531
; Patent No. 6368843
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels Erik E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6368843el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/09/694,531
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 09/198,955
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Clostridium thermocellum
US-09-694-531-11

Alignment Scores:
Pred. No.: 6.97e-13 Length: 1482
```

```
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservativeness: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.18% Indels: 10
DB: 3 Gaps: 2

US-09-917-376-5 (1-88) x US-09-694-531-11 (1-1482)
QY 1 ValSerGlyGlyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGCAATTTGAAGTTGAATCTACACAGCAATCTTCAGATACTACTACTCA 1050
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCCTCAGTTCAAGTTTACTAATACCGAGCAGTGCATTTGATTTGTCCAAACTC 1110
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATTATACAGTACGACGAGCAAGAGATCAGACCTTCTGG---TGTGAC 1167
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCAATTAATCGGCAGTAAACGGCAGTACACGGAATTAATCTCAATGTAAAGGA 1227
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTTGTAAATAATGAGTTCTCTCAACAATAACGACAGACCTACCTTGAA 1278

RESULT 9
US-09-670-141-9
; Sequence 9, Application US/09670141
; Patent No. 6429000
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels Erik K.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Schnorr, Kirk
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
; FILE REFERENCE: 5377.200-US
; CURRENT APPLICATION NUMBER: US/09/670,141
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: 09/198,956
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-670-141-9

Alignment Scores:
Pred. No.: 6.97e-13 Length: 1482
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservativeness: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.18% Indels: 10
DB: 3 Gaps: 2

US-09-917-376-5 (1-88) x US-09-670-141-9 (1-1482)
QY 1 ValSerGlyGlyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGCAATTTGAAGTTGAATCTACACAGCAATCTTCAGATACTACTACTCA 1050
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
```







```
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 4056
; TYPE: DNA
; ORGANISM: Paenibacillus polymyxa
US-09-784-554B-3

Alignment Scores:
Pred. No.: 3,06e-10 Length: 4056
Score: 150.50 Matches: 27
Percent Similarity: 56.4% Conservative: 21
Best Local Similarity: 31.7% Mismatches: 36
Query Match: 32.0% Indels: 1
DB: 4 Gaps: 1

US-09-917-376-5 (1-88) x US-09-784-554B-3 (1-4056)
QY 4 GlyVallyValGlnTyrLysAsnAlaProGlyAspAsnGlnIleLysPro 23
DB 3595 GGAATGGTGCTCCAGTATCGCACGCGGATACAAATGTGAACACATCACTTGACCG 3654

QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
DB 3655 CATTTCCTCAAAATTTAAATAAGGTACATCTCCGTACCGATCAACGAGTTGAAATTCGC 3714

QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63
DB 3715 TACTACTACAGCATCGCGGTGACCGTGAGCAG---ACATTCAACTCGCGACTATGCGGTG 3771

QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
DB 3772 CTGAGCTGCTCGAAGCTGAATCGTAAAGTGGTTAAATGTGATAAAGCTCAACCGGTGCT 3831

QY 84 AspThrTyrLeuGln 88
DB 3832 GATTATTATTGGAA 3846

RESULT 13
US-07-862-588B-6
; Sequence 6, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linnaa
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728

; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1624 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Bacillus lautus
; STRAIN: NCIMB 40250
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 30..1607
; OTHER INFORMATION:
US-07-862-588B-6

Alignment Scores:
Pred. No.: 1.18e-06 Length: 1624
Score: 120.50 Matches: 28
Percent Similarity: 54.12% Conservative: 18
Best Local Similarity: 32.94% Mismatches: 36
Query Match: 25.69% Indels: 3
DB: 2 Gaps: 2

US-09-917-376-5 (1-88) x US-07-862-588B-6 (1-1624)
QY 2 SerGlyGlyVallyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
DB 1323 ACGGGGAACCTTTGTCCTCAATACAAAGTTGGGACACTAGCGCCAGGATAACCAATG 1382

QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
DB 1383 AAGCCTTCCTTTAACATCAAGAACACGGTACAAACCCCTGTTAACTGAGCGGCTCAAG 1442

QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
DB 1443 CTTNNNNNNNNNNNNNAAAAGAC---GGACCTCGGGATATGAGCTGCTCGATCGACTGG 1499

QY 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81
DB 1500 GCGCAAAATCGCCGCAACGAATGTTCTGCTGCGCATTC-----GCTAACTTACCGGGAGT 1553

QY 82 ThrAlaAspThrTyr 86
DB 1554 AATACGGATACTTAC 1568

RESULT 14
US-07-862-588B-5
; Sequence 5, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linnaa
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 5916796 No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
```



```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1775 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Bacillus lautus
; STRAIN: NCIMB 40250
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 30..(1625..1775)
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /partial
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /transl_except= (pos: 1446 .. 1458, aa:
US-07-862-588B-5

Alignment Scores:
Pred. No.: 1.34e-06 Length: 1775
Score: 120.50 Matches: 28
Percent Similarity: 54.12% Conservative: 18
Best Local Similarity: 32.94% Mismatches: 36
Query Match: 25.69% Indels: 3
DB: 2 Gaps: 2

US-09-917-376-5 (1-88) x US-07-862-588B-5 (1-1775)
QY 2 SerGlyGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
DB 1323 ACGGGGAACCTGTGTGTCACAAATGTCGACACTAGCGCCACCGGATAACCAATG 1382
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
DB 1393 AAGCTTCCTTAAATCATCAAGACACGATACCAACCTGTTAACTGAGCGGCTCAAG 1442
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
DB 1443 CTTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1499
QY 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81
DB 1500 GCGCAATTCGGCCGCAACGAATGTTCTGCTGGCATTC-----GCTAACTTTACCGGGAGT 1553
QY 82 ThrAlaAspThrTyr 86
DB 1554 AATCGGATACCTTAC 1568

RESULT 15
US-08-048-164A-1
; Sequence 1, Application US/08048164A
; Patent No. 5496934
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shpiegl, Itai
; APPLICANT: Goldstein, Marc A.
; APPLICANT: Doi, Roy H.
; TITLE OF INVENTION: CELLULOSE BINDING DOMAIN
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/048,164A
; FILING DATE: 14-APR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7809-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 486 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..486
US-08-048-164A-1

Alignment Scores:
Pred. No.: 1.65e-06 Length: 486
Score: 114.00 Matches: 29
Percent Similarity: 52.58% Conservative: 22
Best Local Similarity: 29.90% Mismatches: 34
Query Match: 24.31% Indels: 12
DB: 1 Gaps: 4

US-09-917-376-5 (1-88) x US-08-048-164A-1 (1-486)
QY 2 SerGlyGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
DB 7 ACATCATCAATGTCAGTTGAATTTTCAACTCTAACAAATCAGCACAAACAACTCAATT 66
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
DB 67 ACACCAATAATCAAAATTAATACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAA 126
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
DB 127 GTTAGATATATTATACACAGTGATGTGTACACAGGACAACTTTCTTGG---TGTGACCAT 183
QY 62 AlaAlaMetGlyCysGlyAsn-----IleArgAlaSerPhe 73
DB 184 GCTGTGTCATTTATTAGGAATAGTCTATTGTATACACTAGCAAGTACGACGCAACTTC 243
QY 74 -----GlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
DB 244 GTTAAAGAAACAGCAAGCCCA---ACATCAACCTATGATACATATGTTGAA 291
```



Search completed: October 5, 2005, 22:57:00  
Job time : 47.8869 secs

---



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 5, 2005, 08:30:02 ; Search time 169.52 Seconds  
(without alignments)  
3584.301 Million cell updates/sec

Title: US-09-917-376-5  
Perfect score: 469  
Sequence: 1 VSGGVKQYKKNDSAPDGNQ.....IRASFGSVNPATPTADTYLQ 88

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 7442561 seqs, 3452328358 residues  
Total number of hits satisfying chosen parameters: 14885122

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US09917376/runat\_04102005\_164329\_26954/app\_query.fasta\_1.2524  
-DB=Published Applications NA -OFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09917376 -CGEN 1 1 2178 @runat\_04102005\_164329\_26954  
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCX=100  
-FLOGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-XGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:  
3: /cgn2\_6/ptodata/1/pubpna/US05\_NEW\_PUB.seq:  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:  
6: /cgn2\_6/ptodata/1/pubpna/PTUS\_PUBCOMB.seq:  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:  
13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:  
14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:  
15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:  
16: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:  
17: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:  
18: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq:  
19: /cgn2\_6/ptodata/1/pubpna/US10G\_PUBCOMB.seq:  
20: /cgn2\_6/ptodata/1/pubpna/US10H\_PUBCOMB.seq:  
21: /cgn2\_6/ptodata/1/pubpna/US10I\_PUBCOMB.seq:  
22: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:  
23: /cgn2\_6/ptodata/1/pubpna/US11A\_PUBCOMB.seq:  
24: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq:  
25: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:  
26: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	469	100.0	2289	10	US-09-917-378-2	Sequence 2, Appli
2	466	99.4	3687	10	US-09-917-384-2	Sequence 2, Appli
3	466	99.4	3687	10	US-09-917-383-2	Sequence 2, Appli
4	462	98.5	2869	11	US-09-917-376-2	Sequence 2, Appli
5	462	98.5	2869	15	US-10-155-400-2	Sequence 2, Appli
6	257	54.8	2223	15	US-10-156-761-550	Sequence 550, App
7	257	54.8	9025608	15	US-10-156-761-1	Sequence 1, Appli
8	190	40.5	1527	17	US-10-369-493-46838	Sequence 46838, A
9	188	40.1	1434	19	US-10-466-208-11	Sequence 11, Appl
10	188	40.1	1491	19	US-10-466-208-7	Sequence 7, Appli
11	188	40.1	2510	19	US-10-433-577-34	Sequence 34, Appl
12	165	35.2	501	21	US-10-933-404-1	Sequence 1, Appli
13	165	35.2	1438	17	US-10-372-054-3	Sequence 3, Appli
14	165	35.2	1482	13	US-10-072-152-11	Sequence 11, Appl
15	165	35.2	1482	18	US-10-655-433-11	Sequence 11, Appl
16	157	33.5	1314	9	US-09-974-300-684	Sequence 684, App
17	155.5	33.2	4059	10	US-09-784-554B-1	Sequence 1, Appli
18	155.5	33.2	4059	20	US-10-896-555-1	Sequence 1, Appli
19	150.5	32.1	4056	10	US-09-784-554B-3	Sequence 3, Appli
20	150.5	32.1	4056	20	US-10-896-555-3	Sequence 3, Appli
21	131.5	28.0	3180	22	US-10-886-393A-29	Sequence 29, Appl
22	130.5	27.8	4452	17	US-10-282-122A-15299	Sequence 15299, A
23	126.5	27.0	1303	21	US-10-490-737-1	Sequence 1, Appli
24	126.5	27.0	1747	21	US-10-490-737-2	Sequence 2, Appli
25	126.5	27.0	3489	21	US-10-490-737-3	Sequence 3, Appli
26	114	24.3	768	16	US-10-261-446-5	Sequence 5, Appli
27	114	24.3	768	18	US-10-261-448B-5	Sequence 5, Appli
28	114	24.3	768	21	US-10-782-234-5	Sequence 9, Appli
29	113	24.1	652	20	US-10-748-055-9	Sequence 5, Appli
30	113	24.1	1146	10	US-09-808-898-21	Sequence 21, Appl
31	78.5	16.7	764	20	US-10-653-047-5404	Sequence 5404, Ap
32	78.5	16.7	3447	17	US-10-282-122A-13716	Sequence 13716, A
33	76	16.2	757	19	US-10-767-701-8547	Sequence 8547, Ap
34	76	16.2	96597	18	US-10-052-482-226	Sequence 226, App
35	75.5	16.1	1041	17	US-10-260-238-742	Sequence 742, App
36	75.5	16.1	1041	19	US-10-437-963-76751	Sequence 76751, A
37	75.5	16.1	2091	18	US-10-211-462-222	Sequence 222, App
38	75.5	16.1	2091	20	US-10-723-860-4038	Sequence 4038, Ap
39	75.5	16.1	2091	22	US-10-756-149-4006	Sequence 4006, Ap
40	75.5	16.1	3396	17	US-10-398-458-1	Sequence 1, Appli
41	75.5	16.1	4041	9	US-09-747-835A-21	Sequence 21, Appl
42	75.5	16.1	4041	18	US-10-312-312-21	Sequence 21, Appl
43	75.5	16.1	4353	15	US-10-225-567A-405	Sequence 405, App
44	75.5	16.1	4755	21	US-10-505-486-201	Sequence 201, App
45	75.5	16.1	4914	9	US-09-974-298-188	Sequence 188, App

ALIGNMENTS

RESULT 1  
US-09-917-378-2  
; Sequence 2, Application US/09917378  
; Publication No. US20030119093A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT MANNANASE FROM ACIDOTHERMUS  
; CELLULOXYLICUS  
; FILE REFERENCE: 40197.7US01  
; CURRENT APPLICATION NUMBER: US/09/917,378  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 8







Db 1990 CCGACGGCGGACCTTACCTGCAG 2013  
|||||

RESULT 4

US-09-917-376-2  
; Sequence 2, Application US/09917376  
; Publication No. US20040038334A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOYTICUS  
; FILE REFERENCE: 40197.4US01  
; CURRENT APPLICATION NUMBER: US/09/917,376  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2869  
; TYPE: DNA  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; NAME/KEY: modified\_base  
; LOCATION: (2869)  
; OTHER INFORMATION: a, c, t, g, other or unknown  
US-09-917-376-2

Alignment Scores:  
Pred. No.: 7.09e-55 Length: 2869  
Score: 462.00 Matches: 86  
Percent Similarity: 100.00% Conservative: 2  
Best Local Similarity: 97.73% Mismatches: 0  
Query Match: 98.51% Indels: 0  
DB: 11 Gaps: 0

US-09-917-376-5 (1-88) x US-09-917-376-2 (1-2869)

QY 1 ValSerGlyValValValValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
Db 2605 GTGTGGGTGGGTGAAGGTGAGTATAGATATATGATTCGGCGCGGTGATATCAG 2664  
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrVal 40  
Db 2665 ATCAAGCGGGTTTCAGGTGCGTGAATACCGGGTCGTCTCGGTGGATTGTGCGCGGTG 2724  
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
Db 2725 ACGGTGCGGTACTGGTTTCACCGGGATGGTGGCTCGTGACACTGGTGTACACTGTGAC 2784  
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
Db 2785 TGGCGCGCGATCGGGTGGGATATCCCGCTCGTTCGGTGGTGGTGGTGGTGGTGGTGGTGG 2844  
QY 81 ProThrAlaAspThrTyrLeuGln 88  
Db 2845 CCGACGGCGGACACCTACCTGCAG 2868

RESULT 5

US-10-155-400-2  
; Sequence 2, Application US/10155400  
; Publication No. US20030108988A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOYTICUS  
; FILE REFERENCE: NREL 01-36A  
; CURRENT APPLICATION NUMBER: US/10/155,400  
; CURRENT FILING DATE: 2002-10-22

; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2869  
; TYPE: DNA  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; NAME/KEY: modified\_base  
; LOCATION: (2869)  
; OTHER INFORMATION: a, c, t, g, other or unknown  
US-10-155-400-2

Alignment Scores:  
Pred. No.: 7.09e-55 Length: 2869  
Score: 462.00 Matches: 86  
Percent Similarity: 100.00% Conservative: 2  
Best Local Similarity: 97.73% Mismatches: 0  
Query Match: 98.51% Indels: 0  
DB: 15 Gaps: 0

US-09-917-376-5 (1-88) x US-10-155-400-2 (1-2869)

QY 1 ValSerGlyValValValValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
Db 2605 GTGTGGGTGGGTGAAGGTGAGTATAGATATATGATTCGGCGCGGTGATATCAG 2664  
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrVal 40  
Db 2665 ATCAAGCGGGTTTCAGGTGCGTGAATACCGGGTCGTCTCGGTGGATTGTGCGCGGTG 2724  
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
Db 2725 ACGGTGCGGTACTGGTTTCACCGGGATGGTGGCTCGTGACACTGGTGTACACTGTGAC 2784  
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
Db 2785 TGGCGCGCGATCGGGTGGGATATCCCGCTCGTTCGGTGGTGGTGGTGGTGGTGGTGGTGG 2844  
QY 81 ProThrAlaAspThrTyrLeuGln 88  
Db 2845 CCGACGGCGGACACCTACCTGCAG 2868

RESULT 6

US-10-156-761-550  
; Sequence 550, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 550  
; LENGTH: 2223  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..  
US-10-156-761-550

Alignment Scores:



Pred. No.: 8.85e-26 Length: 2223  
Score: 257.00 Matches: 49  
Percent Similarity: 71.26% Conservative: 13  
Best Local Similarity: 56.32% Mismatches: 25  
Query Match: 54.80% Indels: 0  
DB: 15 Gaps: 0

US-09-917-376-5 (1-88) x US-10-156-761-550 (1-2223)

Qy 2 SerGlyGlyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21  
Db 1771 TCGGGCGGCTTCCAGGTCCTCTCAAGAACACGACTCTCTCGGCCACCGAACGCCCATC 1830  
Qy 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41  
Db 1831 CGGCGAGGCTTCCAGGTCCTCAACCGGAGCGGCTCTCTCGACTGTCCAGGTCACG 1890  
Qy 42 ValArgTyrTrpPheThrArgAspGlyGlySerThrLeuValTyrAsnCysAspTrp 61  
Db 1891 GCCCGCTACTTTCACCGGACAGCGGCTCGGCCACCGTGAACGCTGTGGTGGACTAC 1950  
Qy 62 AlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81  
Db 1951 GCGGCGCTCGGCTGTCCACGCTGAGCTGAAGTGTGTACCCCTGACCAACCGCGTGGCG 2010  
Qy 82 ThrAlaAspThrTyrLeuGln 88  
Db 2011 GGAGCGGAGCTTACTCGAA 2031

## RESULT 7

US-10-156-761-1  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1  
; LENGTH: 9025608  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4187715)  
; OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1

Alignment Scores:  
Pred. No.: 5.62e-21 Length: 9025608  
Score: 257.00 Matches: 49  
Percent Similarity: 71.26% Conservative: 13  
Best Local Similarity: 56.32% Mismatches: 25  
Query Match: 54.80% Indels: 0  
DB: 15 Gaps: 0

US-09-917-376-5 (1-88) x US-10-156-761-1 (1-9025608)

Qy 2 SerGlyGlyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21  
Db 706216 TCGGCGGCTTCCAGGTCCTCTCAAGAACACGACTCTCTCGGCCACCGAACGCCCATC 706275

Qy 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41  
Db 706276 CGGCCAGGCTTCCGATCGTCAACACCGGACGGCTCCCTCGACCTGTCCAAGGTCAAG 706335  
Qy 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61  
Db 706336 GCCCGCTACTTTCACCGGACAGCGGCTCGGCCACCGTGAACGCCCTGTGGTGGACTAC 706395  
Qy 62 AlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81  
Db 706396 GCGGCGGCTGTTCACGCTGAGCTGAAGTGTGTACCCCTGACCAACCGCGGTGGCG 706455  
Qy 82 ThrAlaAspThrTyrLeuGln 88  
Db 706456 GGAGCGGAGCTTACTCGAA 706476

## RESULT 8

US-10-369-493-46838  
; Sequence 46838, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 46838  
; LENGTH: 1527  
; TYPE: DNA  
; ORGANISM: Bacillus subtilis  
US-10-369-493-46838

Alignment Scores:  
Pred. No.: 1.94e-16 Length: 1527  
Score: 190.00 Matches: 35  
Percent Similarity: 65.12% Conservative: 21  
Best Local Similarity: 40.70% Mismatches: 26  
Query Match: 40.51% Indels: 4  
DB: 17 Gaps: 2

US-09-917-376-5 (1-88) x US-10-369-493-46838 (1-1527)

Qy 4 GlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23  
Db 1084 GGTATTTCTGTACAGTACAGCAGCGGGATGGGATGATGAACAGCAACCAATCCGTCG 1143  
Qy 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43  
Db 1144 CAGCTTCAATAAATAAATAACCGCAATACACGGTTGATTAAAGATGTCACCTGCCCGT 1203  
Qy 44 TyrTrpPhe---ThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAla 62  
Db 1204 TACTGTATTAAGCGAATAAACAAGGCCCAAC-----TTTGACTGTGACTAGCG 1254  
Qy 63 AlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThr 82  
Db 1255 CAGATTGGATGGCGCAATGTGACACACAAAGTTTGTGACGTTGTGATTAACCAAGAGT 1314  
Qy 83 AlaAspThrTyrLeuGln 88  
Db 1315 GCAGATACCTATCTGAA 1332

## RESULT 9

US-10-466-208-11



```

; Sequence 11, Application US/10466208
; Publication No. US20040180348A1
; GENERAL INFORMATION:
; APPLICANT: GENOFOCUS Co., Ltd.
; TITLE OF INVENTION: Method for Surface Display of Proteins on Genetic Carriers
; FILE REFERENCE: 3260-13
; CURRENT APPLICATION NUMBER: US/10/466,208
; CURRENT FILING DATE: 2003-07-14
; PRIOR APPLICATION NUMBER: KR2001-2156
; PRIOR FILING DATE: 2001-01-15
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: KopatentIn 1.71
; SEQ ID NO 11
; LENGTH: 1434
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: CMCase gene with additional his encoding sequences
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1431)
US-10-466-208-11

Alignment Scores:
Pred. No.: 3,44e-16 Length: 1434
Score: 188.00 Matches: 34
Percent Similarity: 61.18% Conservative: 18
Best Local Similarity: 40.00% Mismatches: 31
Query Match: 40.09% Indels: 2
DB: 19 Gaps: 1

US-09-917-376-5 (1-88) x US-10-466-208-11 (1-1434)

QY 4 GlyVallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 991 GGTATTTCTGTACAAATACAGACGCGGATGGAGTATGACACGAAACCAATCGTCGC 1050

QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 1051 CAGCTTCAATAAATAAATAACGGCAATACCGGTTGATTTAAAGATGTCACTGCCGT 1110

QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaLa 53
Db 1111 TACTGGTATAACGCGGCAAAACAAAGCGCAAAACGTT-----GACTGTGACTACGCGCAG 1164

QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 1165 CTTGGATCGGCAATGTGACATACAAAGTTTGTGACGTTCATATAACCAAGCAAGTGCA 1224

QY 84 AspThrTyrLeuGln 88
Db 1225 GATACCTATCTGGAA 1239

RESULT 10
US-10-466-208-7
; Sequence 7, Application US/10466208
; Publication No. US20040180348A1
; GENERAL INFORMATION:
; APPLICANT: GENOFOCUS Co., Ltd.
; TITLE OF INVENTION: Method for Surface Display of Proteins on Genetic Carriers
; FILE REFERENCE: 3260-13
; CURRENT APPLICATION NUMBER: US/10/466,208
; CURRENT FILING DATE: 2003-07-14
; PRIOR APPLICATION NUMBER: KR2001-2156
; PRIOR FILING DATE: 2001-01-15
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: KopatentIn 1.71
; SEQ ID NO 7
; LENGTH: 1491
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: CMCase gene with mutated signal sequence to enhance its

```

```

; OTHER INFORMATION: hydrophobicity
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1488)
US-10-466-208-7

Alignment Scores:
Pred. No.: 3,62e-16 Length: 1491
Score: 188.00 Matches: 34
Percent Similarity: 61.18% Conservative: 18
Best Local Similarity: 40.00% Mismatches: 31
Query Match: 40.09% Indels: 2
DB: 19 Gaps: 1

US-09-917-376-5 (1-88) x US-10-466-208-7 (1-1491)

QY 4 GlyVallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 1048 GGTATTTCTGTACAAATACAGACGCGGATGGAGTATGACACGAAACCAATCGTCGC 1107

QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 1108 CAGCTTCAATAAATAAATAACGGCAATACCGGTTGATTTAAAGATGTCACTGCCGT 1167

QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaLa 63
Db 1168 TACTGGTATAACGCGGCAAAACAAAGCGCAAAACGTT-----GACTGTGACTACGCGCAG 1221

QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 1222 CTTGGATCGGCAATGTGACATACAAAGTTTGTGACGTTCATATAACCAAGCAAGTGCA 1281

QY 84 AspThrTyrLeuGln 88
Db 1282 GATACCTATCTGGAA 1296

RESULT 11
US-10-433-577-34
; Sequence 34, Application US/10433577
; Publication No. US20040171065A1
; GENERAL INFORMATION:
; APPLICANT: Genofocus Co., Ltd.
; TITLE OF INVENTION: Method for Expression of Proteins on Spore Surface
; FILE REFERENCE: PCT-Genofocus-1/10/433,577
; CURRENT APPLICATION NUMBER: US/10/433,577
; CURRENT FILING DATE: 2003-06-05
; PRIOR APPLICATION NUMBER: KR2000-74835
; PRIOR FILING DATE: 2000-12-08
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: KopatentIn 1.71
; SEQ ID NO 34
; LENGTH: 2510
; TYPE: DNA
; ORGANISM: Bacillus subtilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (461)..(2491)
US-10-433-577-34

Alignment Scores:
Pred. No.: 7,25e-16 Length: 2510
Score: 188.00 Matches: 34
Percent Similarity: 61.18% Conservative: 18
Best Local Similarity: 40.00% Mismatches: 31
Query Match: 40.09% Indels: 2
DB: 19 Gaps: 1

US-09-917-376-5 (1-88) x US-10-433-577-34 (1-2510)

QY 4 GlyVallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 2051 GGTATTTCTGTACAAATACAGACGCGGATGGAGTATGACACGAAACCAATCGTCGC 2110

```



```
Qy 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 2111 CAGCTTCAATATAAATAACGCAATACCAACGCTTGATTTAAAAAGATGTCACCTCCGCT 2170
Qy 44 TyrTrpPheThrArgAspGlyGlySerThrLeuValThrValAsnCysAspTrpAla 63
Db 2171 TACTGGTATACGCGAATAAACAAGGCGCAAAACGTT-----GACTGTGACTACGCGAG 2224
Qy 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 2225 CTTGATGCGGCATGTGCATACACAGTTTGTGAGTTGTGCATTAACCAACCAAGGTGCA 2284
Qy 84 AspThrTyrLeuGln 88
Db 2285 GATACCTATCTGGAA 2299

RESULT 12
US-10-933-404-1
; Sequence 1, Application US/10933404
; Publication No. US20050118729A1
; GENERAL INFORMATION:
; APPLICANT: MORAG, Ely
; TITLE OF INVENTION: MICROARRAYS OF CELLULOSE BINDING CHIMERIC PROTEINS AND METHODS OF
; FILE REFERENCE: THEREOF
; FILE REFERENCE: 85189-7100
; CURRENT APPLICATION NUMBER: US/10/933,404
; CURRENT FILING DATE: 2004-09-03
; PRIOR APPLICATION NUMBER: PCT/IL03/00177
; PRIOR FILING DATE: 2003-03-06
; PRIOR APPLICATION NUMBER: US 60/362,061
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Clostridium thermocellum
US-10-933-404-1

Alignment Scores:
Pred. No.: 1,62e-13 Length: 501
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.18% Indels: 10
DB: 21 Gaps: 2

US-09-917-376-5 (1-88) x US-10-933-404-1 (1-501)
Qy 1 ValSerGlyGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 13 GTATCAGGCAATTTGAAGTTGAATTTACACAGCAATCTTCAGATACTACTACTCA 72
Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 73 ATCATCTCTCAGTTCAGGTTACTATATACCGAGCAAGTGCATTTGATTTGTCCAAATC 132
Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 133 ACATTGAGATATTATTATACAGTACGAGCAGAAAGATCAGACCTTCTGG---TGTGAC 189
Qy 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 190 CATGCTGCAATTAATCGGCAGTAACGGCAGCTACACGGAATTTACTTCAAAATGTAAAGGA 249
Qy 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 250 ACATTTGTAATAATGAGTTCTCTCAACAAATAACGAGACACCTACCTTGAA 300

RESULT 13
US-10-372-054-3
; Sequence 3, Application US/10372054
; Publication No. US20030203466A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Kauppinen, Markus
; APPLICANT: Schulein, Martin
; APPLICANT: Schnorr, Kirk
; APPLICANT: Andersen, Lene
; APPLICANT: Bjornvad, Mads
; TITLE OF INVENTION: Novel Mannanases
; FILE REFERENCE: 5440.204-US
; CURRENT APPLICATION NUMBER: US/10/372,054
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: US/09/339,159B
; PRIOR FILING DATE: 1993-06-24
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1438
; TYPE: DNA
; ORGANISM: Bacillus sp. I633
US-10-372-054-3

Alignment Scores:
Pred. No.: 6,59e-13 Length: 1438
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.18% Indels: 10
DB: 17 Gaps: 2

US-09-917-376-5 (1-88) x US-10-372-054-3 (1-1438)
Qy 1 ValSerGlyGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 940 GTATCAGGCAATTTGAAGTTGAATTTACACAGCAATCTTCAGATACTACTACTCA 999
Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1000 ATCAATCTCTCAGTTCAGGTTACTATATACCGAGCAGTGCATTTGATTTGTCCAAATC 1059
Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1060 ACATTGAGATATTATTATACAGTACGAGCAGAAAGATCAGACCTTCTGG---TGTGAC 1116
Qy 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1117 CATGCTGCAATTAATCGGCAGTAACGGCAGCTACACGGAATTTACTTCAAAATGTAAAGGA 1176
Qy 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1177 ACATTTGTAATAATGAGTTCTCTCAACAAATAACGAGACACCTACCTTGAA 1227

RESULT 14
US-10-072-152-11
; Sequence 11, Application US/10072152
; Publication No. US20020142438A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. US20020142438A1el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/10/072,152
; CURRENT FILING DATE: 2002-02-07
; PRIOR APPLICATION NUMBER: US/09/198,955
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
```



```
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Clostridium thermocellum
US-10-072-152-11

Alignment Scores:
Pred. No.: 6,86e-13 Length: 1482
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.18% Indels: 10
DB: 13 Gaps: 2

US-09-917-376-5 (1-88) x US-10-072-152-11 (1-1482)
QY 1 ValSerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGCAATTGAGGTTGAATCTACACAGCAATCTTCAGATACCTACTACTCA 1050
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCCTCAGTTCAGGTTACTAATACCGAAGCAGTGCAATGATTGTCCAACTC 1110
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATTATACAGTAGACGACGACAGAAAGATCAGACCTTCTGG---TGTGAC 1167
QY 61 TtpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCAATAATCGGCAGTAAACGCGAGCTACACGGAATTACTTCAATGTAAAGGA 1227
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTGTAAAAATGAGTTCTCTCAACAAATAACGACAGACACCTACTCTTGA 1278

Search completed: October 6, 2005, 05:10:28
Job time : 920.52 secs
```

```
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Clostridium thermocellum
US-10-072-152-11

Alignment Scores:
Pred. No.: 6,86e-13 Length: 1482
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.18% Indels: 10
DB: 13 Gaps: 2

US-09-917-376-5 (1-88) x US-10-072-152-11 (1-1482)
QY 1 ValSerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGCAATTGAGGTTGAATCTACACAGCAATCTTCAGATACCTACTACTCA 1050
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCCTCAGTTCAGGTTACTAATACCGAAGCAGTGCAATGATTGTCCAACTC 1110
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATTATACAGTAGACGACGACAGAAAGATCAGACCTTCTGG---TGTGAC 1167
QY 61 TtpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCAATAATCGGCAGTAAACGCGAGCTACACGGAATTACTTCAATGTAAAGGA 1227
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTGTAAAAATGAGTTCTCTCAACAAATAACGACAGACACCTACTCTTGA 1278

RESULT 15
US-10-655-433-11
; Sequence 11, Application US/10655433
; Publication No. US20040067572A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schuelein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Biorinvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Samme O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. US20040067572A1el Pectate Lyases
; FILE REFERENCE: 5378,200-US
; CURRENT APPLICATION NUMBER: US/10/655,433
; CURRENT FILING DATE: 2003-09-04
; PRIOR APPLICATION NUMBER: US/09/198,955A
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
```